THE RELIABILITY OF LABORATORY PERFORMANCE TESTING

Thesis submitted for the degree of Master of Science (Med) Exercise Science

Elske Jeanne Schabort

The Medical Research Council and The University of Cape Town

Bioenergetics of Exercise Research Unit

Department of Physiology

Sports Science Institute of South Africa

Newlands 7700

Cape Town

South Africa

September 1997

"IF"

"If you can keep your head while all about you
Are losing theirs and blaming it on you,
If you can trust yourself when all men doubt you,
But make allowance for their doubting too;
If you can wait, and not be tired by waiting,
Or being lied about, don't deal in lies,
Or being hated, don't give way to hating,
And yet don't look too good, nor talk too wise:

If you can dream ~ and not make dreams your master;
If you can think ~ and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build'em up with worn-out tools:

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breath a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you,
Except the will that says to them: 'Hold on!'

If you can talk with crowds and keep your virtue,
Or walk with Kings ~ nor lose the common touch,
If neither foes nor loving friends can hurt you,
If all men count with you, but none too much;
If you can fill the unforgiving minute
With sixty seconds' worth of distance run,
Then yours is the Earth, and everything that's in it
And ~ which is more ~ you'll be a Man, my son."

Rudyard Kipling

ACKNOWLEDGEMENTS

I wish to thank and express my sincere appreciation to the following people for their contribution to this thesis:

Dr John Hawley, for his exceptional supervision, confidence in my abilities and for involving me in various projects, always enabling me to gain more knowledge. For the example he sets. For his time, effort and constant advice regarding work and "life". And most of all, for his precious friendship.

Dr Will Hopkins, for the opportunity to have him as co-supervisor, and in particularly his patience and willingness to work over such a long distance. For his highly acknowledged expertise, dedication and the enthusiasm with which he approaches a task.

Professor Tim Noakes, for granting me the opportunity and privilege to further my studies at his highly acclaimed international laboratory. For always supporting and encouraging me to further my goals. For his example and enthusiasm.

Harold Blum and **Iñigo Mujika,** co-authors and co-investigators in these studies, and for sharing their knowledge.

Garry Palmer, for originally suggesting this area of research, for his knowledge and friendship.

My **family** and **friends** for their moral support and motivation.

Glenda Gous, and all the **staff** and **fellow students** of the Sports Science Institute of South Africa, for their support, encouragement and continual friendliness.

To all the subjects who participated in the experiments and without whom the thesis would not have been possible. Thank you for your time and great effort.

Energade (Bromor foods) for their financial assistance and making the experimental work undertaken in this thesis possible.

During my studies I have been fortunate to be supported by research grants and scholarships from a variety of sources. These include The University of Cape Town **Marion Beatrice Waddell scholarship** and **Federation for Research and Development Grant-holder bursary**.

Finally, to the anonymous external examiners for their time, effort and expertise reviewing this thesis.

DECLARATION

I, **Elske Jeanne Schabort**, hereby declare that the work on which this thesis is based is my original work and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the University of Cape Town to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed by candidate

signature removed

Signature

30 September 1997

Date

ABSTRACT

The reproducibility of a measurement in a laboratory test impacts on the power of that test to detect the small, but significant changes in an athlete's performance when determining the influence of a new training or nutritional intervention. Until recently, however, sport scientists have not been concerned with establishing the reliability of many of their testing protocols. Therefore, the purpose of this thesis was to examine the reliability of several laboratory tests of performance and to determine those factors which may impact on the reproducibility of those tests. Possible factors that could contribute to the reliability of a performance test include the type of exercise protocol employed (continuous, intermittent), the equipment on which the subject performs the test, the intensity and duration of the testing protocol, the subject's state of fitness and whether he is familiar with the testing conditions.

In the first study, the reproducibility of performance of distance runners completing a 60-min time trial (TT) on a motor-driven treadmill was determined. Eight trained distance runners (peak oxygen consumption [VO_{2peak}] 66 ± 5 $ml.min^{-1}.kg^{-1}$, mean ± SD) performed the TT on three occasions separated by 7-10 days. Throughout each TT the runners controlled the speed of the treadmill and could view their current speed and elapsed time, but not the elapsed or final distance. On the basis of heart rate (HR) data, it is estimated that the subjects ran at an average intensity equivalent to 80-83% of VO_{2peak}. The distance run in 60-min did not vary substantially between trials (16.2 ± 1.4 km, 15.9 ± 1.4 km, and 16.1 ± 1.2 km for TT_{1-3} respectively, $p = 0.5$). The coefficient of variation (CV) for individual runners was 2.7% (95% confidence interval [CI] of 1.8 to 4.0%) and the test-retest reliability expressed as an intraclass correlation coefficient was 0.90 (95%CI 0.72 to 0.98). Reproducibility of

performance in this test was therefore acceptable. However, higher reproducibility is required for experimental studies aimed at detecting the smallest worthwhile changes in performance with realistic sample sizes.

In an attempt to determine the reliability of a performance test over a race distance which subjects are familiar with, eight well-trained rowers (VO_{2peak}, 61 ± 5 ml.kg.$^{-1}$min^{-1}; peak power output [PPO], 346 ± 35 W) performed a 2,000-m TT on a Concept II™ rowing ergometer on three occasions separated by three-day intervals. During each trial, the rowers knew the elapsed distance and were told their split time for each successive 500-m, but they were not informed of their final time for any trial until the completion of the study. The final times for the three TTs showed a small learning effect (6:57 ± 0:15, 6:54 ± 0:14 and 6:51 ± 0:14 min:sec, for TT_1, TT_2 and TT_3, respectively), a very small CV (0.6%, 95%CI 0.4 to 1.0) and a very high intraclass correlation coefficient (0.97, 95%CI 0.91 to 0.99). Such high reliability makes this test suitable for monitoring the performance of rowers and for investigating interventions that affect performance in short, high-intensity endurance events.

The final study examined the reliability of a prolonged cycling protocol undertaken in the laboratory which included bouts of exercise at different workloads such as occurs in many race situations. Eight endurance trained cyclists (VO_{2peak}, 64.8 ± 5.7 ml.kg.$^{-1}$min^{-1}, PPO, 411 ± 43 W) performed a 100-km TT on three occasions, separated by 5-7 days. Subjects were free to regulate the intensity at which they exercised throughout the rides. Four 1-km and four 4-km sprints were included in the TT during which the subjects were asked to cycle as hard as possible. They also had to complete the 100-km in as fast a time possible. During the trials subjects were only allowed to see their HR and the distance covered, and were not

informed of their overall time until completion of the final TT. The final times of the three TTs showed a small improvement from TT_1 to TT_3 (151:52 ± 10:36, 148:24 ± 9:42 and 147:24 ± 8:48 min:sec, for TT_1, TT_2 and TT_3, respectively). The CV for individual cyclists was 1.7% (95%CI 1.1 to 2.5) and the test-retest reliability expressed as an intraclass correlation coefficient was 0.93 (95%CI 0.79 to 0.98). The results of this final study indicate that a laboratory performance test with intermittent bouts of high intensity exercise, undertaken on the subject's own bicycle, is highly reproducible. Laboratory tests in which subjects are allowed to freely choose their effort rather than a predetermined workload being imposed on them, have proved to be more reliable than previous exercise tests to exhaustion at a fixed workload.

In conclusion, the results of the experiments conducted in this thesis suggest that when evaluating an experimental intervention, sport scientists should employ laboratory performance tests which have high reproducibility. Such reliability is improved when athletes are allowed to utilise their own specialised equipment and tests are conducted over the same distance or for approximately the same time as the athlete's specialised event. Also, athletes must be allowed to self-select their own pacing strategy rather than a fixed workload being imposed on them. Such high reliability is essential if sport scientists are to detect the small, but worthwhile changes in performance.

PUBLICATIONS

Schabort EJ, Hopkins WG, Hawley JA. (1997) Reproducibility of self-paced treadmill performance of trained endurance runners. *International Journal of Sports Medicine* (in press).

Schabort EJ, Hopkins WG, Hawley JA and Blum H. (1997) High reliability of performance of well-trained rowers on a rowing ergometer. *Journal of Sport Sciences* (in review).

Schabort EJ, Hawley JA, Hopkins WG and Mujika I. (1997) A new, reliable laboratory performance test for endurance trained cyclists. *Medicine and Science in Sport and Exercise* (in review).

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION AND AIMS OF THIS THESIS

Over the past 50 years, there have been progressive improvements in world records in all athletic events. Some of these performance improvements have been associated with better training and nutritional practices, technological advances, better equipment, as well as input from sport scientists. Indeed, the establishment of sports institutes around the world has assisted exercise scientists in collecting physiological, biomechanical and nutritional data on elite competitors with the aim of assisting them to optimise their training programmes. Evidence of the benefit of such institutions can be seen from those nations which have established sports centres in their countries and have been rewarded with improved athletic success. As an example, the Australian Institute of Sport in Canberra was established in 1980 and since then the performances of that nation's athletes have improved significantly. At the Montreal Olympic Games in 1976, Australia won only five medals, with no gold medals. However, at the 1992 Barcelona Olympics Games, they won 27 medals which included seven gold medals. Four years later at the Atlanta games, Australia won a total of 41 medals, of which nine were gold.

In addition to optimising performance through training and other interventions, results from competitive athletic events have also provided information on how physiological variables interact to improve performance. Therefore, sport scientists have focused on identifying and improving the specific factors that can assist an athlete to reach their ultimate potential. To assist in these tasks, numerous laboratory and field tests have been developed to assess an athlete's strengths and weaknesses relative to their specialised event and to monitor their responses to various training interventions.

Until recently, sport scientists have not been concerned with the reliability or validity of many of their testing measures, and valuable work has been undertaken without the scientist taking into account the reproducibility or validity of their performance measures. Such information is necessary if scientists are to be able to detect the small changes in performance which often separate the superior athlete from the rest of the field.

With regard to the reliability of performance tests, the most commonly employed laboratory protocols have typically employed the time to "exhaustion" at some fixed, arbitrary workload as a measure of exercise performance. However, such protocols have poor reproducibility. Also, these tests are not related to the competitive situation in which athletes compete over a set distance or for a certain duration of time, rather than exercising until exhaustion.

The reliability of performance in this thesis was examined during three different laboratory exercise tests that could be used in future research to determine factors that affect performance of both endurance athletes, as well as the performance of athletes competing in short, high-intensity events. The tests that were used span over a wide range of intensity and duration. To access athletes who were experienced in competing over each duration, a different mode of exercise was used for each duration: rowing for short, maximal intensity tests, cycling for the endurance-trained athletes and running as a high-intensity, endurance test. With regard to the cycling test, short bouts of high-intensity exercise was included to make the test more closely resemble competitive endurance cycling.

The first study, described in Chapter Three, examined the reliability of a performance test conducted on a motorised treadmill. In this study, endurance trained runners had to run as far as possible in a given time. They were free to select the pace at which they exercised, as well as change the intensity as they progressed through the trial. Asking athletes to perform for a fixed, known duration should result in a more reliable outcome measure than an open-ended task to exhaustion.

The second study, described in Chapter Four, assessed the reliability of a short, high intensity exercise test conducted on a rowing ergometer. Well-trained rowers performed several performance time-trials (TTs) over their standard racing distance of 2,000-m on a rowing ergometer. They were free to select the pace at which they exercised. Asking an athlete to produce a maximal effort over their normal competitive distance might result in better reproducibility compared to exercising for a fixed time.

The final study (Chapter Five) examined the reliability of a prolonged endurance cycling performance test. The trials were conducted on the subjects' own bicycles. Multiple sprints were included in the TT so that the subjects exercised at various intensities to resemble race conditions more closely. A combination of familiar equipment and an exercise task which mimics race conditions in the field should further enhance the reliability of performance testing.

CHAPTER TWO

LITERATURE REVIEW

THE RELIABILITY OF LABORATORY PERFORMANCE TESTING

2.1 The purpose of physiological testing for athletes

Laboratory testing of athletes to identify athletic potential dates back to the 1920's when studies to determine the maximal oxygen uptake (VO_{2max}) of runners and swimmers were first undertaken (Herbst, 1928; Liljestrand and Stenstrom, 1920). These studies established that oxygen uptake (VO_2) rose with increases in running and swimming speed. They also indicated that trained runners had the highest VO_{2max} values when tested during running, rather than swimming. These, and subsequent investigators (Astrand, 1952; Robinson, 1937) were the first to identify the important role of VO_{2max} for successful performance in endurance events. In addition, they highlighted the importance of testing athletes in their specific events for test results to be valid.

In the past years, a large number of laboratory tests have been devised. Regardless of the testing protocol employed, there are common objectives for any physiological testing programme. These are summarised below:

i. To aid the sport scientist in the construction of sport-specific physiological profiles of individual athletes or teams,
ii. To indicate the athlete's strengths and weaknesses relevant to their event,
iii. To provide baseline data for training prescription,
iv. To monitor any physiological and performance changes in an athlete or team,
v. To provide feedback to coaches so that they can evaluate the effectiveness of their training interventions,
vi. To aid in talent identification, and

vii. To provide insight into the condition of the athlete and the type and magnitude of physiological changes that might be feasible to optimise performance.

2.2 Criteria for an athlete testing programme

Before a sport scientist can conduct any laboratory test, a number of methodological criteria must be fulfilled. These are summarised below (Hawley, 1997a; McDougall et al., 1991):

i. Tests must be reliable, therefore, measurements must be performed consistently and the outcome must be reproducible. A test is of little value if its reliability is not sufficiently high to reflect the slight changes that might have occurred in the athlete over a training period or after a specific intervention has been introduced,
ii. Tests must be valid: they must measure what they claim to measure,
iii. The variables that are tested must be relevant to that sport: the physiological components that have application to their particular sport must be tested,
iv. Tests must be sensitive enough to detect small changes in the athlete's state of fitness and performance relevant to their chosen event,
v. The tests must be sport-specific. For the test results to be of practical significance, the mode of exercise must be specific to the sport. Despite the fact that a test might be highly reliable, the validity declines as the testing protocol becomes more removed from the variable that is supposed to be tested,
vi. Test administration must be rigidly controlled. Once test variables are selected, they must be conducted consistently at all times. This includes the standardisation of testing protocols, laboratory conditions, equipment and calibration procedures. In addition, variables such as the nutritional and training status of the athlete, time of day when the

test takes place and other interventions such as sleep, illness or injury, must also be taken into consideration,

vii. Testing must be repeated at regular intervals. If the main objective of testing is to monitor the effectiveness of a particular training programme or other intervention, then it is required that these tests are repeated at regular intervals.

Recently, there has been an increasing demand to determine the effects of various training and nutritional strategies as athletes seek to enhance their performance. When evaluating a potential ergogenic aid, it is important that certain testing criteria should be satisfied. The most important of these criteria is probably the reliability of a test or measure. However, it is important to acknowledge that often there is a trade-off between strict scientific standards (such as reliability and validity) and the practicality of a sports specific testing protocol. This review will focus on the reliability or reproducibility of laboratory tests which determine performance.

2.3 Reliability of laboratory tests

Reliability refers to the reproducibility of a test or measurement. The reliability or reproducibility of a test is a crucial issue for the sport scientist in deciding on the utility of a test. Reproducibility is important to determine whether the test will be sensitive enough to detect the small changes in an athlete's physiological status resulting from a new training or nutritional strategy.

The variability of a test is commonly expressed as a coefficient of variation (CV). CV is the standard deviation (SD) of a measure divided by the mean of that measure ($CV = SD/\bar{X} \times 100$). In some studies, the variability has been reported as a correlation coefficient. This is most common in those investigations which have examined only two trials and expressed the data in terms of a test-retest correlation (Bar-Or, 1981; Dotan and Bar-Or, 1980). The CV is useful for researchers who are concerned with detecting changes of a variable when the magnitude of change is interpreted in relation to the variability within a subject from test to test. The correlation coefficient, on the other hand, takes into account variation between subjects, so it is more useful when the magnitude of change is interpreted in terms of variation between subjects.

The variability of a given test is influenced by two main sources: technological variability and biological variability. Technological variability is due mainly to instrumental errors, differences in calibration of equipment, changes in environmental conditions, and other technical or human errors (Coggan and Costill, 1984). For example, with regard to instrumental error, rapid response gas analysers have errors of only ± 0.01% of full scale when correctly calibrated (McConnell, 1988), whereas the CV for the calibration-gas concentrations can be as low as ± 0.03% (Weber and Janicki, 1986). Modern automated metabolic units have an accuracy of measurement ventilation of ± 4.0% when volume is measured (McConnell, 1988) with a 2.0% deviation from linearity when flow is integrated over time (Harrison et al., 1980). For most laboratory tests, the technological error is relatively constant and does not increase or decrease in proportion to the magnitude of the outcome variable being tested (Henry, 1959), indicating that a greater percentage of total variation of a performance test will be due to biological variability as the CV increases.

On the other hand, biological variability is due to the inherent biological fluctuation of an organism, and refers to the day-to-day variation in energy metabolism. Under standardised laboratory conditions, up to 90% of the total variation in an individual's response to a maximal incremental exercise test to exhaustion (VO_{2max}), can be accounted for by biological variability (Coggan and Costill, 1984; Katch et al., 1982). Both technological and biological variability must be taken into account by sport scientists when interpreting the results of laboratory performance tests. In this thesis, biological variability will be discussed.

One of the first studies to determine the extent of technological variability on laboratory performance, was that of Taylor (1944). He studied 31 normally active students who each performed two incremental treadmill tests to exhaustion, separated by three days. The protocol consisted of a four-min walk at 108 $m.min^{-1}$ at a 5% gradient, followed by a four-min rest, after which time the progressive test was started. This involved running at a constant speed of 162 $m.min^{-1}$ and at an initial gradient of 5% which was elevated by 1% per min until exhaustion. Results from this experiment revealed that errors in measurement contributed less than 1% of the total variance in the outcome measure (VO_{2max}). In contrast, in a review of the literature, Shephard (1984) has estimated that a greater percentage (at least 10%) of total variability can be attributed to technical error.

To assess the biological and technological variability of several "anaerobic" tests of performance, Coggan and Costill (1984) examined four maximal, exhaustive protocols lasting either 30 sec or 60 sec (Table 2.1). The mean CV was 5.4% for the mean power output and 6.2% for peak power output (PPO) attained during the 30-sec and 60-sec tests, respectively. For a timed ride to exhaustion at 125% of VO_{2max}, the CV was similar (5.3%). The

technological error for this test was determined from 20 repeated calibrations of all the equipment used, with the CV taken to represent net technological error. Only 10%-30% of the variability could be accounted for by technological error, whereas biological variability contributed the rest of the variability. The data from this study confirm that biological variability accounted for almost all of the variability in anaerobic test measurements (Coggan and Costill, 1984).

The variability of measures of maximal aerobic power are similar to those reported for anaerobic power. Katch et al. (1982) studied five trained females and one trained male who each performed between eight and 20 VO_{2max} tests on a treadmill over a two-four week period. The total variability for VO_{2max} amounted to ± 5.6%, with biological variability contributing >90% and technological variability <10% (Katch et al., 1982). Boileau et al. (1977) have estimated the intra-individual day-to-day variability in VO_{2max} to range from 4% to 6% in moderately fit individuals undertaking regular exercise, whereas Shephard (1984) reported a 10% day-to-day variation in the physical condition of subjects. Katch et al. (1982) further speculated that the biological variation would be larger in untrained subjects due to increased variation in both the transport and the extraction of oxygen at the cellular level, and that a further factor contributing to biological variation could be motivation.

In contrast to high-intensity tests, Armstrong and Costill (1985) examined the technological and biological variability during submaximal running and cycling in five runners and five cyclists, each group performing both the running and cycling trials. Cycling trials were conducted on an electrically braked bicycle ergometer at workloads of 100, 150 and 250 W, whereas the running trials were conducted at treadmill speeds of 170, 200 and 230 $m.min^{-1}$.

Each steady-state cycling and running workload was maintained for five min. They determined that the CV for the day-to-day variability in submaximal VO_2 ($l.min^{-1}$) was 4% for both cycling and running trials. Subject day-to-day variability accounted ~90% of the total day-to-day variation (Armstrong and Costill, 1985; Table 2.1). These workers speculated that the training state of the subjects, their previous experiences with the cycling ergometer and treadmill or alterations in the economy of movements may have accounted for the day-to-day variations (Armstrong and Costill, 1985).

In summary, the extent of variability for tests of both aerobic and anaerobic power due to technological error is small and usually insignificant when compared to the contribution from biological variation (Coggan and Costill, 1984; Henry, 1959; Katch et al., 1982; Taylor, 1944).

The following Table displays the biological and technological variability of several laboratory tests of physiological function.

Table 2.1 The biological and technological variability of some frequently used laboratory tests.

Test measure	Total variability (%)	Biological (%)	Technological (%)	Reference
Peak torque (30 sec)	6.7	76.1	23.9	Coggan and Costill, 1984
Peak torque (60 sec)	5.6	71.4	28.6	Coggan and Costill, 1984
Mean power (30 sec)	5.4	68.5	31.5	Coggan and Costill, 1984
Mean power (60 sec)	5.4	71.4	28.6	Coggan and Costill, 1984
% Fatigue (30 sec)	10.3	84.5	15.5	Coggan and Costill, 1984
% Fatigue (60 sec)	7.5	78.7	21.3	Coggan and Costill, 1984
Time to exhaustion	5.3	90.0	10.0	Coggan and Costill, 1984
VO_{2max} ($l.min^{-1}$)	2.4	29.0	1.0	Taylor et al., 1944
VO_{2max} ($l.min^{-1}$)	5.6	92.7	7.3	Katch et al., 1982
VO_{2max} ($ml.kg.^{-1}min^{-1}$)	5.6	92.6	7.4	Katch et al., 1982
Cycling VO_2 @ 100 W	4.7	4.3	0.4	Armstrong and Costill, 1985
150 W	4.7	4.3	0.4	
200 W	3.9	3.5	0.4	
Running VO_2 @ 170 $m.min^{-1}$	4.1	3.6	0.4	Armstrong and Costill, 1985
200 $m.min^{-1}$	4.5	4.1	0.4	
230 $m.min^{-1}$	2.9	2.5	0.4	

2.3.1 Reliability of laboratory tests of exercise capacity

Laboratory tests of exercise capacity are those tests which claim to measure the physiological factors which have a strong relationship with successful performance of a specified event in the field. These should be differentiated from laboratory tests of performance *per se* which will be discussed in Section 2.3.2.

The most frequently used tests for determining the aerobic fitness status of athletes from a variety of sports is the test for VO_{2max}. Indeed, judging by the frequency with which this test is discussed in scientific and lay publications, it is clear that most athletes still believe VO_{2max} to be the best predictor of athletic potential (Noakes, 1988). For the accurate determination of athletic potential, one should consider the reliability of this and other performance tests and whether they are sensitive enough to determine small changes in performance capacity. Tables 2.2 and 2.3 describe the reliability of laboratory tests of exercise capacity in cycling and running and show that there is a wide range in the variability of VO_{2max} for both well-trained and untrained but "active" subjects. The CV for this measure ranges from 1.0% (Taylor et al., 1955) to as much as 9.1% (Graham and Andrew, 1973). The practical implication of such a wide and unsystematic variability in VO_{2max} makes it difficult to determine whether a change in this measure is due to a true change in aerobic capacity, or whether it is merely due to technological and/or biological variability. For example, if one was to study the effects of a conditioning programme on aerobic capacity in a group of five subjects with a mean VO_{2max} value of 46 $ml.kg.^{-1}min^{-1}$, the mean VO_{2max} value after conditioning will have to exceed 51 $ml.kg.^{-1}min^{-1}$ before a true effect could reliably be detected (Kyle et al., 1989). It is worth noting that the variability in VO_{2max} is of the same

order of magnitude whether this measure is expressed in relative ($ml.kg.^{-1}min^{-1}$) or absolute ($l.min^{-1}$) terms.

A possible reason for differences in the variability of measurements of maximal exercise capacity, might be because untrained subjects may not be sufficiently motivated or willing to push themselves to a point which produces a true "maximum" effort. Indeed, as previously mentioned, Katch et al. (1982) suggested that variation might be greater in untrained subjects where motivation could play a large role in contributing to such a large variability.

The data summarised in Tables 2.2 to 2.5 suggests that fitter, well-trained subjects might be better able to reproduce their performance when compared to untrained, less active subjects. However, in contrast to most reports, Kyle et al. (1989) found that the variability in repeated measures of peak oxygen uptake (VO_{2peak}) was not different across different training level groups. They studied VO_{2peak} in five highly trained, seven moderately trained and five untrained males performing three maximal treadmill tests to exhaustion. No effect of different fitness levels of the subjects was observed for the variability in VO_{2peak}. The within-subject CV for VO_{2peak} ($ml.kg.^{-1}min^{-1}$) was 4.9%, whereas the CV for treadmill time was 3.9%.

Tests for the measurement of anaerobic power have shown to have a higher reliability when compared to tests for the measurement of PPO. The test-retest reliability of the Wingate anaerobic test have ranged from 0.89 to 0.98, but are typically in the order of 0.94 (for review, see Bar-Or, 1987). For cyclists (Table 2.2), measures of both mean and PPO during

the Wingate anaerobic test can be determined with a CV of only 0.96% (Evans and Quinney, 1981).

In running, critical power has been used an exercise test for monitoring an athlete's adaptation to training. Critical power is typically defined as the power output which can be maintained indefinitely or without exhaustion by the athlete. The protocol for determination of critical power was developed from that originally described by Hughson et al. (1984). Briefly, subjects were required to run on the treadmill at six different speeds, ranging from 17-25 $km.hr^{-1}$ until they were unable to maintain the set pace. The slowest speed for each subject equalled the speed run for their fastest 5 km time, plus 2 $km.hr^{-1}$. The remaining five speeds were calculated to cause the subjects to become fatigued in 2-12 min. Least-squares analysis was used to fit an exponential decay to the relationship between the running speed (y) versus time to exhaustion (x). Critical power was calculated as the running speed (y) coinciding with the asymptote or C parameter of the relationship $y = A.e^{(-Bx)} + C$ (Kolbe et al., 1995). This test has shown to be highly reliable for runners, with a test-retest correlation coefficient of 0.99 (Kolbe et al., 1994). It has been suggested that when performing treadmill tests, one might expect the measurements to be less affected by familiarisation when compared to cycling, since it involves a more common skill than pedalling (Graham and Andrew, 1973).

There have been few studies examining the reliability of sports other than running or cycling. Henry et al. (1995) studied 13 collegiate rowers on two occasions performing 30-sec bouts of maximal upper body exercise on the Concept II rowing ergometer, and during a modified Wingate test for upper body power output. Both peak and mean power output determined on

the Concept II rowing ergometer, as well as during the Wingate tests, were found to be highly reliable. The correlation coefficient for PPO was 0.94 and 0.98 and for mean power output 0.96 and 0.98, for the Concept II and Wingate tests, respectively.

For swimming, VO_{2max} tests have reported to be highly reliable, with correlation coefficients ranging from 0.92 (Léger et al., 1980) to 0.98 (Costill et al., 1985). Costill et al. (1985) determined the reliability of VO_{2max} measurements using two different protocols: i) tethered breaststroke swimming and ii) a 20-sec gas sample taken immediately after a 366-m front crawl swim. For the first protocol, 39 (25 male and 14 female) swimmers were studied during seven min of tethered breaststroke swimming and with a 80-sec recovery period from exercise. One-min expired gas samples were collected during the 6^{th} and 7^{th} min of the swim and every 20 sec during the 80-sec recovery. The correlation between VO_2 during exercise and the first 20-sec recovery period, was 0.98. In the second protocol, 52 swimmers from various age groups were studied. Each swimmer performed two 400-yd freestyle trials at maximal effort. The swimmers were instructed to take a breath approximately one stroke before the finish of the 366-m swim and subsequently exhale the breath into a breathing mask which sealed over their face approximately one sec after finishing. The reliability coefficient for this test-retest was 0.97 (refer to Table 2.5).

Table 2.2 Reliability of laboratory tests of exercise capacity - *Cycling*

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
12 active adults	Wingate anaerobic test (mean and peak power output)	2	-	0.96	Evans and Quinney, 1981
19 active males	PPO: a) 4.41 joule.rev.$^{-1}$kg^{-1} BW b) 5.59 joule.rev.$^{-1}$kg^{-1} BW	2 tests separated by 1-3 days	-	a) peak: 0.93 mean: 0.93 b) peak: 0.91 mean: 0.93	Patton et al., 1985
8 endurance trained cyclists	Incremental cycling test to exhaustion : PPO (W)	3 tests performed over 4-5 week period	1.14 ± 0.65 (0.62-2.5)	-	Lindsay et al., 1996
10 physically active men	Incremental cycling test to exhaustion to: a) PPO (W) b) VO_{2max} (l.min^{-1})	1 test.week^{-1} performed over 9-12 month period	a) 4.99 (2.95-6.83) b) 7.89 (4.2-11.35)	-	Kuipers et al., 1985

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
4 untrained subjects	VO_{2max}	tests performed over 4 month period	4.7 (3.5-5.7)	-	Wyndham et al., 1959
21 12-yr old boys	Incremental cycling test to exhaustion: a) VO_{2max} ($l.min^{-1}$) b) VO_{2max} ($ml.kg.^{-1}min^{-1}$) c) Performance time	2 tests separated by 1 week	-	a) 0.95 b) 0.88 c) 0.91	Boileau et al., 1977

CV, coefficient of variation; r, correlation coefficient; PPO, peak power output; BW, body weight; $V0_{2max}$, maximal oxygen uptake. Values are mean ± SD.

Table 2.3 Reliability of laboratory tests of exercise capacity - *Running*

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
a) 12 male sprint runners b) 13 physically active men	Maximal anaerobic running test (maximal power: $ml.kg.^{-1}min^{-1}$)	2	a) 1.58 b) 2.75	a) 0.95 b) 0.92	Nummela, 1996
18 male distance runners	VO_{2max} (performance time)	2 tests separated by 1 week	± 1.96	0.95	Farrell et al., 1979
5 athletes (4 trained females and 1 male)	Incremental running test to exhaustion: a) VO_{2max} ($l.min^{-1}$ and $ml.kg.^{-1}min^{-1}$) b) performance time	8-20 tests performed over 2-4 week period	a) 5.6 (3.7-7.3) b) 5.95 (4.1-7.8)	-	Katch et al., 1982
21 12-yr old boys	Incremental running test to exhaustion: a) VO_{2max} ($l.min^{-1}$) b) VO_{2max}($ml.kg.^{-1}min^{-1}$) c) performance time	2 tests separated by 1 week	-	a) 0.97 b) 0.87 c) 0.76	Boileau et al., 1977

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
41 college women	Incremental running test to exhaustion: Balke protocol a) $l.min^{-1}$ b) $ml.kg.^{-1}min^{-1}$	2 tests performed over 3-5 day period	-	a) 0.95 b) 0.91	McArdle et al., 1972
31 active students	Incremental running test to exhaustion: a) performance time b) VO_{2max} ($l.min^{-1}$)	2 tests performed within 3 days	b) 2.4	a) 0.95 b) 0.70	Taylor, 1944
17 trained and untrained male runners	Incremental running test to exhaustion: a) performance time b) VO_{2max} ($ml.min^{-1}$) c) VO_{2max} ($ml.kg.^{-1}min^{-1}$)	3 tests performed within 1 month	a) 3.9 b) 4.6 c) 4.9	-	Kyle et al., 1989
6 physically active students	a) VO_{2max} ($ml.kg.^{-1}min^{-1}$) b) VO_{2max} ($l.min^{-1}$)	5-7 tests performed over 3½ month period	a) 9.1 (5.6-12.1) b) 9.0 (6.4-11.2)	-	Graham and Andrew, 1973
a) 1 Cross-country skier b) 1 Oarsman	VO_{2max} ($l.min^{-1}$)	1 $test.week^{-1}$ performed over a total of 17 weeks	a) 6.8 b) 5.1	-	Wright, 1978

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
28 untrained subjects	VO_{2max} ($l.min^{-1}$)	2 tests performed over 3-5 day period	0.7-7.6	0.95	Taylor et al., 1955
36 untrained male and female subjects	VO_{2max} using backward-extrapolation from O_2-recovery curve	2	-	0.92	Léger et al., 1980
8 long distance runners	Critical power	2 tests performed within 1 week	-	0.99	Kolbe et al., 1995

CV, coefficient of variation; r, correlation coefficient; $V0_{2max}$, maximal oxygen uptake.

Table 2.4 Reliability of laboratory tests of exercise capacity - *Rowing*

Subjects	Test protocol	Number of trials	CV (%)	r	Reference
13 college rowers	Wingate anaerobic test: a) peak power (W) b) mean power (W) Concept II ergometer: c) peak power (W) d) mean power (W)	2 tests performed within 6 days	-	a) 0.98 b) 0.98 c) 0.94 d) 0.96	Henry et al., 1995

CV, coefficient of variation; r, correlation coefficient

Table 2.5 Reliability of laboratory tests of exercise capacity - *Swimming*

Subjects	Test protocol	Number of trials	CV (%)	r	Reference
a) 25 male and 14 female competitive swimmers b) 37 male and 15 female competitive swimmers	VO_{2max} ($ml.kg.^{-1}min^{-1}$) a) tethered breaststroke swimming b) 365.8 m maximal freestyle swim	2 tests separated by 24 hrs	-	a) 0.98 b) 0.97	Costill et al., 1985

CV, coefficient of variation; r, correlation coefficient; $V0_{2max}$, maximal oxygen uptake.

2.3.2 Reliability of laboratory tests of performance

The use of continuous, fixed-load, submaximal exercise testing protocols has been the most common method of testing performance in the laboratory. In these tests, athletes are requested to exercise at a fixed, submaximal workload until exhaustion, which is typically defined as the inability to sustain the given power output or speed of movement. Such testing protocols, however, have limitations. Firstly, in sports such as cycling, there are a continuous change in exercise intensity due to changes in pace, hill-climbs, drafting or "break-away" sprints when the cyclists are free to regulate their power output (Palmer et al., 1994). Indeed, when athletes can chose their own pacing strategy, such as during competition, they consistently achieve greater physiological responses than those responses achieved during exercise bouts which are fixed at a constant intensity and performed in the laboratory where the subjects are not influenced by external factors (Foster et al., 1993a; Padilla et al., 1996; Palmer et al., 1994).

Secondly, tests to volitional fatigue or "exhaustion" have been shown to have poor reproducibility. Tables 2.6 and 2.7 summarise the data from studies which have determined the reliability of laboratory tests of "performance" for cycling and running. It has been observed that in both cycling (Table 2.6) and running (Table 2.7), large variability exists when exercise time to exhaustion protocols at fixed, submaximal workloads in both well-trained and untrained individuals are examined, with the CV being as high as 56% (Krebs and Powers, 1989). The high variability in these tests might be related to the fact that in open-ended trials, subjects don't know when they will be finishing and therefore they find it difficult to prepare themselves mentally for an unknown period of time. Also, as the time of the performance test increases, it is possible that factors other than the energy demand, such

as diet, hydration and motivation, could influence the results when subjects complete their exercise at different end-points in these longer performance tests (McLellan et al., 1995).

Various studies have been conducted on cycling exercise to exhaustion. Krebs and Powers (1989) studied 10 male volunteers who cycled to exhaustion at 80% of VO_{2max} on two separate occasions. The CV for this task ranged from 5.2% to 56%, with a mean CV of 20%. It could be argued that the poor CV was the result of a large learning effect between the two trials when subjects with various abilities are studied. In order to address this question, McLellan et al. (1995) studied 15 untrained cyclists exercising at the same intensity (80% VO_{2max}) on five different occasions. The exercise time on this task ranged from 14-18 min. They (McLellan et al., 1995) observed CVs ranging from 2.8% to 31% and reported a minimal learning effect from trial-to-trial.

It could also be expected that the fitness level of subjects influences the CV and that the reason for the large variation found in these studies (Krebs and Powers, 1989; McLellan et al., 1995) was because untrained individuals were tested. However, when Jeukendrup et al. (1996) studied nine well-trained cyclists and triathletes who cycled to exhaustion at 75% of VO_{2max} (a workload which exhausted the subjects in ~1 hr), the resulting CV was also high. In that study, the CV was 27%, ranging from 17% to 40%.

Billat et al. (1994) studied the reproducibility of high intensity running time to exhaustion at maximal aerobic speed on eight male long-distance runners. Average running time was six-seven min. Subjects performed two tests and the average difference in running time between

the two tests was 44-sec or approximately 10% of total time to exhaustion. For three of the subjects this difference exceeded 60-sec. Within-subject variability amounted to 25%.

In contrast to the large variability in exercise tasks to exhaustion lasting between 15-60 min, supra-maximal "time to exhaustion" tasks lasting <90 sec have shown to have good reproducibility. Lindsay et al. (1996) determined the reproducibility of high intensity cycling time to exhaustion at 150% of PPO in eight competitive cyclists, repeated on three occasions. The rides averaged ~67 sec. The CV for this task was only 1.7% ± 1.3%.

In contrast to these tests in which athletes are required to exercise to exhaustion at a fixed, submaximal workload, performance tests in which athletes are asked to perform a certain amount of work or cover a set distance in the shortest time possible, or complete as much work as possible within a certain period of time (Tables 2.6 and 2.7), have shown to be much more reliable than constant-load tests.

Jeukendrup et al. (1996) studied two groups of ten, well-trained subjects who completed either as much work as possible in 60 min, or performed a 15-min TT after 45 min of constant submaximal cycling at 70% of their PPO. Each subject performed one of the two experimental protocols on six different occasions. They reported CVs of 3.4% and 3.5%, respectively and concluded that there was no learning effect in trained subjects, implying that it is not necessary to perform an extra learning trial when subjects are accustomed to the laboratory procedures or have participated in previous laboratory tests. Similarly, Bishop (1997) reported a CV of 2.7% for 20 trained female subjects also performing two 60-min TTs in which they had to generate as much power as possible. It should be noted that even though

the subjects were defined as "trained cyclists", their average VO_{2peak} was only 47.4 ± 7.2 $ml.kg.^{-1}min^{-1}$ which would be considered "active" as opposed to "trained".

High reproducibilities in cycling time-trials over 20 km and 40 km have also been reported when trained subjects are able to ride their own bicycles on air-braked cycle ergometers (Hawley et al., 1997b; Lindsay et al., 1996; Palmer et al., 1996; Table 2.6), as well as on a conventional cycle ergometer (Hickey et al., 1992; Table 2.6). A problem with the standard laboratory ergometers is that experienced cyclists have difficulty in assuming their normal riding position during testing, therefore air-braked cycle simulators have been used by trained cyclists during trials which allow them to ride their own bike in the laboratory.

Palmer et al. (1996) studied six well-trained, competitive cyclists who each undertook three 20-km and three 40-km TTs on an air-braked cycle ergometer (Kingcycle Ltd, High Wycombe, Bucks, U.K.). The time taken for the laboratory simulated TTs were highly reproducible, with CVs of 1.1 ± 0.9% and 1.0 ± 0.5% (mean ± SD) for the 20-km and 40-km, respectively. Using equipment which the cyclists are accustomed to and which is used during competition could be a reason for the high reliability found when using these ergometers. Alternatively, it may well be that when subjects are free to regulate their own effort and employ their own pacing strategy, they are better able to reproduce performance than when a fixed workload is imposed on them.

Others have also studied the reliability of cycling tests of performance. Hickey et al. (1992) assessed the reproducibility of TT laboratory cycling performance at three different workloads in eight well-trained cyclists. The trials were completed on an isokinetic cycle

ergometer (Cybex MET 100) which enabled subjects to self-pace each trial and adjust cadence in 5 rev.min^{-1} increments during the course of the trials. Each subject performed three trials as fast as possible which consisted of 1600, 200 and 14 kJ (equivalent to approximately 40, 5 and 0.5 miles). The mean CV of each of the three trials were ± 1.01%, ± 0.95% and ± 2.43% for the 40, 5 and 0.5 mile trials, respectively.

Low CVs for the reliability of running and cycling economy were reported by Armstrong and Costill (1985) during submaximal running and cycling. They studied five cyclists and five runners, each group performing both four running and four cycling trials over an eight to nine day period. Cycling trials were conducted at 100, 150 and 250 W, whereas the running trials were set at velocities of 170, 200 and 230 m.min^{-1}. Each workload or running speed was maintained for five min. The CV for VO_2 (l.min^{-1}) during the three cycling workloads (n = 10), were 4.7%, 4.7% and 3.9% for 100, 150 and 230 W, respectively. The mean variation for the three workloads was 4.4%. For the running trials, the CV for each of the three running speeds were 4.1%, 4.5% and 2.9% at 170, 200 and 230 m.min^{-1}, respectively, with the mean variation being 3.8%. They (Armstrong and Costill, 1985) speculated that the training state of the subjects, as well as previous experiences with laboratory testing, may have influenced the CV.

In summary, several factors may explain the differences in measures of reliability. Psychological factors can significantly influence endurance performance and effort sensations during TT protocols (Hickey et al., 1992; Jeukendrup et al., 1996; Palmer et al., 1996), such as a decrease in performance time during the last ride of a trial which may be due to the cyclist's knowledge that it is his final ride (Hickey et al., 1992).

The characteristics of the population studied will also influence the outcome of the reliability of the performance test. Results of performances among trained subjects cannot be extrapolated to populations that are untrained. With regard to familiarisation of subjects, there have been contrasting findings. Whereas some workers have suggested that physically fit subjects who are experienced in laboratory testing and familiar with the nature of the experimental protocol will minimise the effect of any training or learning effects (Coggan and Costill, 1984; Hickey et al., 1992; Jeukendrup et al., 1996), others (Graham and Andrew, 1973; Kyle et al., 1989) concluded that their results were not affected by familiarisation associated with repeated testing or the fitness level of their subjects.

In conclusion, a review of the contemporary literature suggests that trained athletes are better able to optimise their effort when performing TTs (Hickey et al., 1992; Jeukendrup et al., 1996; Palmer et al., 1996), incremental tests to exhaustion (Boileau et al., 1977; Kyle et al., 1989; McArdle, 1972; Taylor, 1944) or running to exhaustion at speeds eliciting VO_{2max} (Billat et al., 1994) when compared to exercise tests to exhaustion at a constant workload (Jeukendrup et al., 1996; Krebs and Powers, 1989; McLellan et al., 1995). The use of simulated TT tasks are preferable to laboratory tests of exercise capacity in that they provide a further measure of performance and compare better to race conditions.

Table 2.6 Reliability of laboratory tests of performance - *Cycling*

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
3 groups of 9 endurance trained cyclists and runners	a) Anaerobic power during 30 sec sprint (W) b) Anaerobic power during 60 sec sprint (W) c) Peak torque during 30 sec sprint ($n.m^{-1}$) d) Peak torque during 60 sec sprint ($n.m^{-1}$) e) % Fatigue during 30 sec sprint f) % Fatigue during 60 sec sprint g) Timed ride to exhaustion at 125% VO_{2max}	4 tests performed within 4 week period	a) 5.4 b) 5.4 c) 6.7 d) 5.6 e) 10.3 f) 7.5 g) 5.3	a-f) 0.89-0.97	Coggan and Costill, 1984
8 endurance trained cyclists	Timed ride to exhaustion at 150% PPO (60-80 sec)	3 tests performed over 4-5 week period	1.7 ± 1.3 (0.0-3.69)	-	Lindsay et al., 1996

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
15 untrained male subjects	Timed ride to exhaustion at 80% VO_{2max} (15-20 min)	5 tests, each separated by at least 72 hrs	17.3 (2.8-31.4)	-	McLellan et al., 1995
10 untrained male subjects	Timed ride to exhaustion at 80% VO_{2max}	2 tests separated by 1 week	20.3 (5.2-55.9)	0.51	Krebs and Powers, 1989
8 moderately active, non-cyclists	Timed tide to exhaustion at 105% of an intensity corresponding to each subject's lactate minimum	2 tests separated by 4 weeks	17.82	-	Caine and McConnell, 1995
a) 9 well-cyclists and triathletes b) 10 well-cyclists and triathletes c) 10 well-cyclists and triathletes	a) Timed ride to exhaustion at 75% VO_{2max} (~1 hr) b) Cycling 45 min at 70% VO_{2max}, followed by 15 min TT c) Time to complete a set amount of work (~1 hr)	6 tests, each separated by 3-7 days	a) 26.6 (17.4-39.5) b) 3.49 (1.2-5.8) c) 3.35 (0.8-5.8)	-	Jeukendrup et al., 1996
20 trained female cyclists	Power output generated in 1-hr TT	2 tests separated by 1 week	2.7%	0.97	Bishop D, 1997

Subjects	Test protocol	Number of trials	CV (%) (range)	r	Reference
8 well-trained male cyclists	i) Time: a) 1600 kJ (~40 miles, ~105 min) b) 200 kJ (~5 miles, ~12 min) c) 14 kJ (~0.5 miles, ~60 sec) ii) Power output: a) 1600 kJ (~40 miles) b) 200 kJ (~5 miles) c) 14 kJ (~0.5 miles)	4 tests, each separated by at least 72 hrs	ia) 1.01 ib) 0.95 ic) 2.43 iia) 1.25 iib)1.53 iic) 3.09	-	Hickey et al., 1992
two groups of 6 highly trained cyclists	Time-trials: a) 20 km (~27 min) b) 40 km (~56 min)	3 tests, each separated by at least 72 hrs	a) 1.1 ± 0.9 b) 1.0 ± 0.5	-	Palmer et al., 1996
8 endurance trained male cyclists	40 km Time-trial (~56 min)	3 tests performed over 4-5 week period	0.89 ± 0.45 (0.24-1.6)	-	Lindsay et al., 1996
5 cyclists	5 min submaximal work at a) 100 W b) 150 W c) 250 W	4 tests performed over 8-9 day period	a) 4.67 b) 4.73 c) 3.86	-	Armstrong and Costill, 1985

CV, coefficient of variation; r, correlation coefficient; PPO, peak power output; $V0_{2max}$, maximal oxygen uptake. Values are mean ± SD.

Table 2.7 Reliability of laboratory tests of performance - *Running*

Subjects	Test protocol	Number of trials	CV (%)	r	Reference
8 long distance runners	Time to exhaustion at maximal aerobic speed (~6-7 min)	2 tests separated by 1 week	26.5	0.86	Billat et al., 1994
5 runners	Submaximal treadmill running @ a) 170 $m.min^{-1}$ b) 200 $m.min^{-1}$ c) 230 $m.min^{-1}$	4 tests performed over 8-9 day period	a) 4.05 b) 4.47 c) 2.89	-	Armstrong and Costill, 1985

CV, coefficient of variation; r, correlation coefficient.

2.4 Summary and Conclusions

To assist sports scientists identify and improve the specific factors which can assist athletes reach their physiological potential, a variety of laboratory tests have been developed. Many of these tests, however, have poor reliability. The reproducibility of a test protocol is a crucial issue for the sports scientist in deciding on the utility of a test. It impacts on the statistical power of a test to determine whether it will be able to detect the small, but worthwhile changes in an athlete's physiological ability which will influence his/her performance.

To date, the most commonly used laboratory procedure to test "performance" have required the athlete to exercise to exhaustion at some arbitrary, fixed submaximal workload or speed. Such protocols, however, have very poor reproducibility. More to the point, athletes rarely, if ever, train or race at a constant workload until volitional fatigue.

Recently, more sports specific laboratory tests have been devised. These allow the athlete to monitor their own pace or effort and are often completed over distances which athletes are used to during competition. Not surprisingly, the reproducibility of these protocols is much better. There are still relatively few studies which have been conducted to systematically determine the optimal laboratory test for sports scientists to measure performance. Indeed, the characteristics of the ideal laboratory test protocol remains to be established.

CHAPTER THREE

REPRODUCIBILITY OF SELF-PACED TREADMILL PERFORMANCE OF TRAINED ENDURANCE RUNNERS

Abstract

The reproducibility of performance in a laboratory test impacts on the power of that test to detect changes of performance in experiments. The purpose of this study was to determine the reproducibility of performance in distance runners completing a 60-min TT on a motor-driven treadmill. Eight trained distance runners (age 27 ± 7 yr, VO_{2peak} 66 ± 5 $ml.kg.^{-1}min^{-1}$, mean ± SD) performed the TT on three occasions separated by 7-10 days. Throughout each TT the runners controlled the speed of the treadmill and could view current speed and elapsed time, but they did not know elapsed or final distance. On the basis of heart rate (HR) data, it was estimated that the subjects ran at an average intensity equivalent to 80-83% of VO_{2peak}. The distance run in 60 min did not vary substantially between trials (16.2 ± 1.4 km, 15.9 ± 1.4 km, and 16.1 ± 1.2 km for TT_{1-3} respectively, $p = 0.5$). The CV for individual runners was 2.7% (95%CI 1.8 to 4.0%) and the test-retest reliability, expressed as an intraclass correlation coefficient, was 0.90 (95%CI 0.72 to 0.98). Reproducibility of performance in this test was therefore acceptable. However, higher reproducibilities are required for experimental studies aimed at detecting the smallest worthwhile changes in performance with realistic sample sizes.

3.1 Introduction

The reproducibility of a performance test when it is administered on two or more occasions to the same subjects is a crucial issue in deciding on the utility of a test. Reproducibility impacts on the power of a test to detect performance changes in research on treatments that may affect performance. Reproducibility is usually expressed as a CV - the within-subject variation expressed as a percent of the subject's mean.

One of the common varieties of endurance tests require the athlete to exercise at a fixed workload until they become exhausted; the measure of performance is then simply the duration of the test. The reproducibility of performance in such constant-load tests on a cycle ergometer is surprisingly poor: CVs of 17-27% have been reported (Jeukendrup et al., 1996; Krebs and Powers 1989; McLellan et al., 1995). Variability in performance of this magnitude would make it difficult for researchers using these tests to detect changes in performance in the order of 1%. Such a seemingly small change of performance is of great interest to athletes, because it appears to be typical of the difference in time that separates the top athletes in an endurance event.

Reproducibility is markedly better in the other varieties of endurance tests. Hickey et al. (1992) asked cyclists to perform several fixed amounts of work as quickly as possible. For the workload that required a mean time of 105 min to perform, the CV for performance time was 1.0%; the same CV was obtained for an amount of work that was performed in a mean time of 12 min. Palmer et al. (1996) observed similar CVs when they asked cyclists to complete simulated 20- and 40-km TTs on a cycle ergometer as quickly as possible (mean times of 27 min and 56 min, respectively). A somewhat poorer CV of 3.4% was obtained

when Jeukendrup et al. (1996) used a slightly different protocol in which a different amount of work was selected for each cyclist in an attempt to make all performance times close to 60 min. These authors also found a CV of 3.5% when cyclists were instructed to perform as much work as possible in 15 min following 45 min of work at a submaximal intensity (70% PPO). Bishop (1997) reported a CV of 2.7% for female subjects generating as much work as possible during a 60-min TT.

It is evident that research on reproducibility of endurance performance has thus far focused on cycling, but researchers who are interested in optimising performance of distance runners also need a test of endurance running that can be administered with confidence in a laboratory setting. The only relevant research on such tests is a study by Billat et al. (1994), who measured exercise time to exhaustion at a speed corresponding to each runner's VO_{2max}). The CV calculated for the time to exhaustion from their data (see Methods) was 25%, for a mean performance time of six to seven min. It is likely that longer endurance tests of this kind for runners would produce similar or even worse reproducibility. Therefore, to investigate a test of running, modelled on the most successful tests of cycling, a 60-min TT was chosen as the exercise protocol in which the subjects ran as far as possible on a treadmill by setting their own pace throughout the test.

3.2 Methods

3.2.1 Subjects

Eight endurance-trained runners, who competed regularly in local races, participated in this study after giving written informed consent. The mean (± SD) age, mass, VO_{2peak} and peak HR (HR_{peak}) of the eight subjects under investigation was 27 ± 7 yrs, 75 ± 12 kg, 66 ± 5

ml.kg.$^{-1}$min^{-1} and 193 ± 7 min^{-1}, respectively. At the time of the investigation subjects were running five to six times per week, for a total average distance of 75 km (range 40-140 km.wk^{-1}).

3.2.2 Preliminary testing

For the maximal test, the subject warmed up for ~10 min on a treadmill (Powerjog, Sport Engineering Limited, Birmingham, England) at a self-selected intensity. After a five-min rest, the subject remounted the treadmill and the speed was increased to 12 km.hr^{-1}. This speed was maintained for 60 sec, after which it was increased by 1 km.hr^{-1}.min^{-1} until volitional fatigue. During the maximal test subjects wore a mask covering the nose and mouth; the expired air passed through an on-line computer system attached to an Oxycon Alpha automated gas analyser (Mijnhardt, The Netherlands) for the determination of oxygen consumption. Before each test, the gas analyser was calibrated with a Hans Rudolph 5530, 3 L syringe and an online CO_2:N_2 gas mixture of know composition. Analyser outputs were processed by an IBM computer, which calculated minute ventilation, oxygen consumption and carbon dioxide production using conventional equations (Jones, 1982). Each subject's VO_{2peak} was taken as the highest O_2 uptake measured during any 60-sec period of the test (Noakes et al., 1990).

HR during the maximal test and the subsequent TTs was recorded with a Polar Sport Tester HR-monitor (Polar Electro OY, Kempele, Finland).

3.2.3 Time trials

Each subject completed three TTs separated by 7-10 days. Each TT was scheduled so that it did not interfere with the subject's normal training and racing programme. Each TT for a given subject was conducted at the same time of the day, and subjects were instructed to maintain the same diet and training regimen for 48 hr prior to a TT. In addition, the subjects were asked not to perform any hard training on the day prior to a TT. During all testing, laboratory conditions were standardised at a temperature of 20°C and a relative humidity of 55%. Accuracy of the speed of the treadmill was verified before and after the study. The treadmill was placed in the same position so that subjects had the same view during each trial, and a fan was positioned to cool the subjects during their trials.

Before each trial, subjects were fitted with a HR-monitor and allowed a brief (5-8 min) self-paced warm up on the treadmill. After a short rest, during which some of the subjects performed stretching exercises, they were requested to "run as far as possible in 60 min" and therefore instructed to adjust the speed of the treadmill accordingly during each trial. The gradient of the treadmill for all tests was 0%. Subjects consumed water ad libitum during each TT and the only feedback given to subjects was their elapsed time and current speed. Subjects were not informed of the final distance they covered in each TT until the completion of the study.

No metabolic (respiratory gases, blood lactate) monitoring was undertaken during the 60-min trials since such an intervention would interfere with the subject's performance.

3.2.4 Statistical analyses

Measures of reproducibility for the distance run in one hr were derived by two-way analysis of variance (ANOVA); subject and TT identities were the two effects, and the natural logarithm of the distance was used as the dependent variable. Log transformation was performed to minimise heteroscedasticity (non-normal distribution of error in the ANOVA), and to permit statistically elegant derivation of the mean CV. Reliability expressed as an intraclass correlation coefficient was calculated as (F -1)/(F + k - 1), where F was the F-ratio for the subject term and k (= 3) was the number of trials. The p-value for the F-ratio of the TT term represented a test of significance of differences in the mean times for the three trials. A mean CV was derived from the root mean square error (RMSE) in the analysis by the following transformation: $CV = 100(e^{RMSE} - 1) \approx 100(RMSE)$. The 95%CI for the intraclass correlation coefficient and CV were calculated by the methods of McGraw and Wong (1996) and Tate and Klett (1959), respectively. All measures of centrality and spread are presented as mean ± SD. CVs for individual runners were calculated by dividing each runner's SD by his mean for the three trials.

Previous researchers in this field have derived the overall CV for a test by averaging the CVs of individual subjects. This method produces a biased estimate of about 0.9 of the correct value. For comparison with the CV of the present study, unbiased estimates of published CVs were obtained by taking the square root of the average of the square of the CVs of individual subjects. A CV was also calculated by the ANOVA method for the data of Billat et al. (1994).

3.3 Results

Table 3.1 shows the physiological responses attained during the three TTs. The overall average speed for the three TTs was 15.9 ± 1.3 $km.hr^{-1}$. The average HR maintained for the duration of the three trials was 168 ± 6 min^{-1}. On the basis of HR and oxygen consumptions measured in the maximal test, we estimate that individual subjects ran at an average of 80-83% of VO_{2peak} for the duration of their TTs (Hawley and Noakes, 1992).

Table 3.2 shows the total distance covered by each subject during each TT, together with their mean distance ± SD and CV for the three runs. The CV for the total distance covered during the TT was 2.7% (95%CI 1.8 to 4.0). The reliability, expressed as an intraclass correlation coefficient, was 0.90 (95%CI 0.72 to 0.98). There was a small increase in reproducibility calculated for TT_2 and TT_3 alone (CV = 2.6%, correlation = 0.91) in comparison with that of TT_1 and TT_2 alone (CV = 3.0%, correlation = 0.88). The differences between mean distances covered in the three trials were not statistically significant (p = 0.5) and were insubstantial relative to the between-subject SDs.

Fig. 3.1A illustrates the average speed of the eight subjects. The average speed increased from 14.8 $km.hr^{-1}$ after 5 min to 16.2 $km.hr^{-1}$ after the first 20-30 min of exercise; the speed then stabilised until 10-15 min from the end, after which it increased progressively from 16.5 to 17.6 $km.hr^{-1}$ at the end of the 60 min. The average HR response of the subjects are shown in Fig. 3.1B. The HR of the eight subjects rose rapidly from ~158 to 165 min^{-1} during the first 15 min of exercise, after which the HR gradually increased to reach 177 min^{-1} by the end of the runs.

Table 3.1 Physiological responses measured during the three 60-min time-trials.

	Distance (km)	Average Speed ($km.hr^{-1}$)	Average HR (min^{-1})
Trial 1	16.16 ± 1.40	16.0 ± 1.3	170 ± 6
(range)	(13.4 - 17.7)	(13.6 - 17.5)	(165 - 181)
Trial 2	15.94 ± 1.37	15.8 ± 1.3	167 ± 6
(range)	(13.3 - 17.8)	(13.2 - 17.6)	(160 - 176)
Trial 3	16.14 ± 1.24	16.0 ± 1.4	168 ± 7
(range)	(13.3 - 17.3)	(13.2 - 17.7)	(158 - 177)
Mean	16.08	15.9	168
SD	1.3	1.3	6

HR, heart rate. Values are mean ± SD for the three time-trials (n = 8).

Table 3.2 Distances (km) run by individual subjects in the three 60-min time-trials.

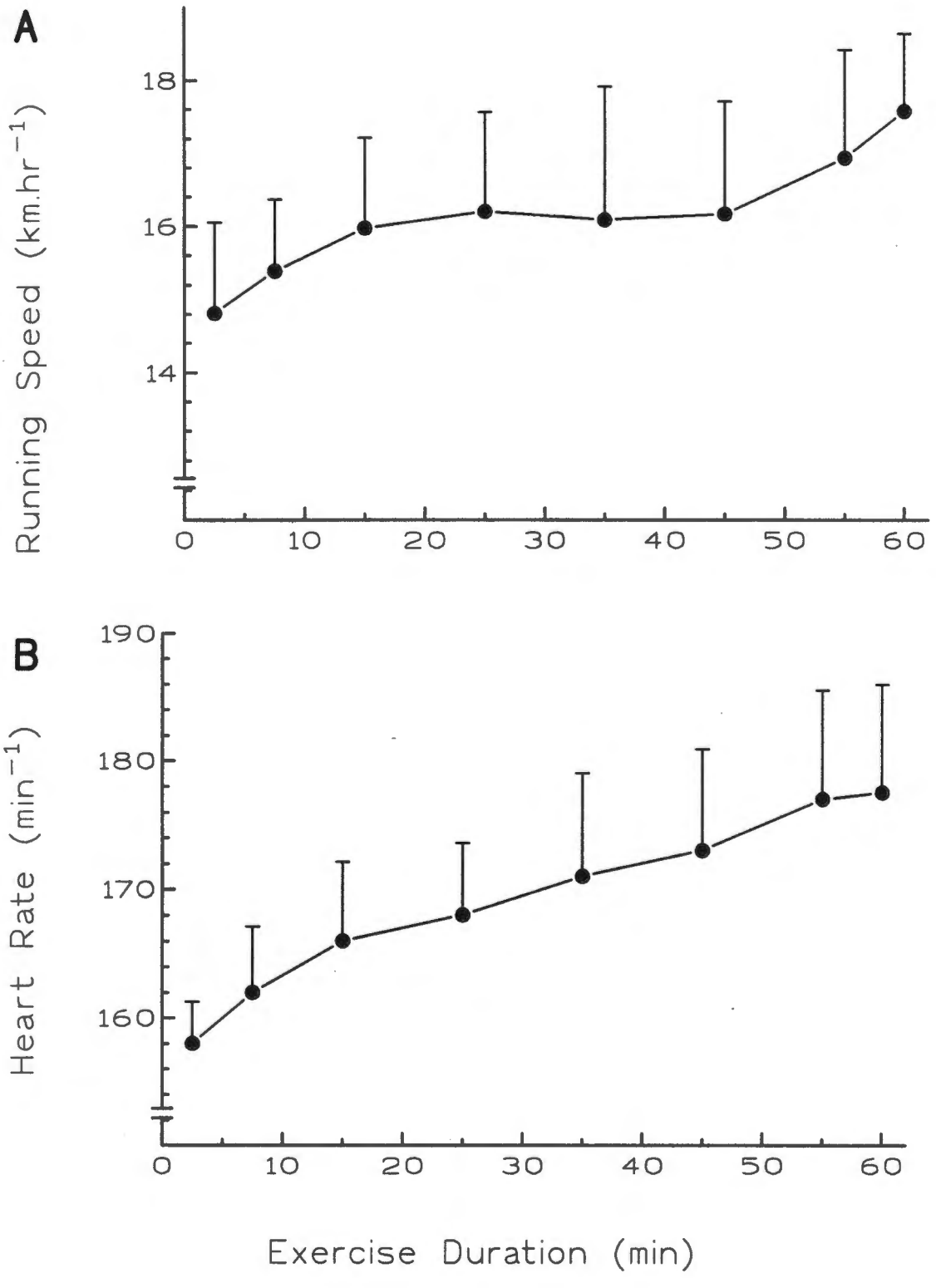
Subject	1	2	3	4	5	6	7	8	Mean	SD
Trial 1	14.9	16.6	17.0	17.3	17.7	13.4	16.0	16.4	16.16	1.39
Trial 2	15.5	15.0	16.6	16.7	17.8	13.3	16.2	16.5	15.94	1.37
Trial 3	15.9	16.3	16.3	17.3	17.2	13.3	16.5	16.4	16.14	1.24
Mean	15.43	15.97	16.66	17.08	17.54	13.32	16.22	16.41		
SD	0.52	0.90	0.33	0.30	0.31	0.08	0.24	0.05		
CV (%)	3.3	5.6	2.0	1.8	1.8	0.63	1.5	0.3	2.7%	

CV, coefficient of variation. Values are mean ± SD for the three time-trials (n = 8).

Figure 3.1A Mean running speed for the three 60-min time-trials. With the exception of the 60-min data points, which are instantaneous measurements, the data shown are averages of 5- to 10-min periods centred on the time point.

Figure 3.1B Mean heart rate for the three 60-min time-trials. The data shown are instantaneous measurements taken at 5- and 10-min periods during the time-trials.

All values are mean ± SD.


A
Running Speed (km.hr^{-1})
18
16
14
0
10
20
30
40
50
60
B
Heart Rate (min^{-1})
190
180
170
160
0
10
20
30
40
50
60
Exercise Duration (min)

3.4 Discussion

Research into the effect of drugs, nutritional substances, or training methods on athletic performance usually involves the assessment of performance in a physical test that simulates the demands of the athlete's sport. Physical tests of endurance performance require the athlete to attempt to maximise work or power while performing prolonged moderate- to high-intensity exercise. Several varieties of endurance tests can be employed, and recently researchers have begun to investigate which type of test is best (Bishop, 1997; Hickey et al., 1992; Jeukendrup et al., 1996; Krebs and Powers, 1989; McLellan et al., 1995; Palmer et al., 1996).

The CV observed for performance in the 60-min running tests fall towards the low end of the range of CVs that have been found for other endurance tests. The confidence interval of the CV overlaps those found by Jeukendrup et al. (1996) for trained cyclists performing either a pre-set amount of work as quickly as possible or as much work as possible in 15 min after a previous bout of fatiguing exercise (3.6-3.8% after correction). Hence, the running test utilised in the current study should be regarded as having reproducibility similar to that of these tests. The running test is clearly superior to the constant-load tests to exhaustion investigated by others (Billat et al., 1994; Jeukendrup et al., 1996; Krebs and Powers, 1989) which produced CVs of 10-27%. On the other hand, the confidence interval falls well short of the CVs of the other tests (1.1-1.2% after correction), in which cyclists performed a fixed amount of work as quickly as possible (Hickey et al., 1992; Palmer et al., 1996). The running test is therefore almost certainly not as reproducible as these cycling tests.

Several factors could be responsible for the differences in reproducibility of endurance tests. Psychological factors such as motivation and boredom are probably the main reasons for the poor reproducibility of the open-ended, constant-load tests (Palmer et al., 1996), whereas athletes evidently can regulate their power output much more reliably when they have an end-point of time, distance, or work to focus on. The type of ergometer used may explain some of the variation in reproducibility between the different cycle tests (Firth, 1981), and it is possible that treadmill running is inherently less reliable than cycling on an ergometer. Differences in competitive experience of the athletes may explain some of the differences between studies, because more experienced athletes are almost certainly better able to judge the pace or effort required for an extended period of exercise. On the other hand, familiarity with the ergometer may be less important, if one considers that, for runners with no previous experience of treadmill running, there was only a small increase in reproducibility after the first trial and no substantial differences in mean performance between the three trials. Finally, details of the protocol almost certainly contribute to the reproducibility of performance. Indeed, it might be possible to enhance reproducibility by giving the athletes more feedback about their progress during the test. For example, information about their speed and elapsed distance might produce a flatter speed profile than which was observed here, which should optimise the pace of the athlete's performance (Foster et al., 1994), and presumably thereby reduce variability. However, further research will be required to test this hypothesis.

Is a CV in the range of 2-4% adequate for research on factors that might affect athletic endurance performance? Definitely not, for the following reason. A change in performance of 1% would win or lose a race for a top athlete, because it represents about 15 sec in a half-

hour race such as a 10-km run. But with a CV of 2%, such a change could be detected only with an unusually high sample size. For example, if a study is designed with the usual type I error rate (5% - the chance of declaring a false positive) and the usual power (80% - the chance of getting statistical significance for a true change in performance of 1%), then standard power calculations show that about 250 athletes would need to be studied: 125 in a control group and 125 in an intervention group (Cohen, 1988). Such large sample sizes are unheard of in experimental sport research. The sample size becomes a more realistic but still daunting 32 in each group when the CV is 1%. It is clear that researchers of sport performance should make every effort to improve the reproducibility of their performance tests and increase their sample sizes if they are to avoid missing small but worthwhile effects on the performance of competitive athletes.

In conclusion, the reproducibility of performance of the treadmill test described in the current investigation was acceptable, but below that observed recently for endurance cycling tests of the same duration.

CHAPTER FOUR

HIGH RELIABILITY OF PERFORMANCE OF WELL-TRAINED ROWERS ON A ROWING ERGOMETER

Abstract

Performance tests with high reliability are needed to investigate factors that result in small, but worthwhile changes in performance. Such reliability is important in all performance tests because sport scientists frequently employ maximal ergometer tests to assess changes in the physiological status of athletes and for coaches, as a tool for team selection. Therefore, the purpose of this study was to determine the reliability of a short, high-intensity 2,000-m TT on a Concept II™ rowing ergometer. Eight well-trained rowers (VO_{2peak} 61 ± 5 ml.kg.$^{-1}$min^{-1}; PPO 346 ± 35 W) performed the TT on three occasions separated by three days. During each trial, the rowers knew their elapsed distance and were told their split time for each successive 500 m, but they were not informed of their final time for any trial until the completion of the study. The rowers performed the tests at 92 ± 5% of PPO and 95 ± 1% of HR_{peak} (mean ± SD). Peak lactate concentration measured after the second TT was 19.4 ± 2.0 mmol.l^{-1}. The final times for the three TTs showed a small learning effect (6:56.5 ± 0:15, 6:54.2 ± 0:14 and 6:51.0 ± 0:14 min:sec for TT_1, TT_2 and TT_3, respectively), a very small CV (0.6%, 95%CI 0.4 to 1.0) and a very high intraclass correlation coefficient (0.97, 95%CI 0.91 to 0.99). Such high reliability makes this test highly suitable for the investigation of interventions that affect performance in short, high-intensity events.

4.1 Introduction

In their attempts to determine the effect of experimental interventions on athletic performance, sport scientists have employed a variety of laboratory and field tests. Such tests need to be highly reliable to detect the smallest worthwhile changes in performance with realistic sample sizes (Hopkins, 1997). The tests should also simulate the physiological demands of the athlete's sport to increase the likelihood that outcomes in the tests apply to performance in real events. Tests meeting these requirements are likely to be useful not only to sport scientists but also to coaches or team selectors who want to track changes in the fitness of individual athletes.

It is clear from recent research that there is a wide variation in the reliability of performance tests. The protocol of the test appears to have the greatest effect on reliability: constant-load tests, in which the athlete exercises to exhaustion at a fixed workload, have indicated to be less reliable than tests in which the athlete attempts to maximise power over a set time or distance (Bishop, 1997; Hickey et al., 1992; Jeukendrup et al., 1996; Krebs and Powers, 1989; McLellan et al., 1995; Palmer et al., 1996). The ergometer, mode of exercise, and duration of exercise are other factors likely to affect reliability (Firth, 1981; Hickey et al., 1992; McLellan et al., 1995). These factors also impact on the extent to which the test simulates the demands of a competitive event.

In the sport of rowing, athletes frequently perform laboratory-based tests on ergometers. In a widely used test that simulates a 2,000-m rowing event, work done on the ergometer is converted to an equivalent distance travelled, and the rower is instructed to cover the distance as quickly as possible. This 2,000-m "all-out" rowing test is often used instead of a

progressive incremental test to exhaustion to evaluate physiological performance in rowers. The Concept II™ rowing ergometer (Morrisville, Vermont, USA) is a popular choice for this laboratory test. Athletes and coaches appear to be satisfied that the ergometer and test protocol reproduce the physical demands of the real event, but there are no published data on the reliability of performance tests on this ergometer. This chapter now provides such data.

4.2 Methods

4.2.1 Subjects

Eight trained rowers who were all members of a high-school first-eight squad participated in this investigation. Prior to any testing, subjects were fully informed of the nature of the investigation, after which they gave their written informed consent. At the time of the investigation, subjects were training seven sessions per week, which included ~80 $km.wk^{-1}$ of on-water training and additional off-water resistance workouts.

4.2.2 Preliminary testing

On their first visit to the laboratory, subjects performed a progressive incremental test to exhaustion on a Concept II™ rowing ergometer for the determination of VO_{2peak} and PPO.

Before commencing the maximal test, subjects warmed up on the rowing ergometer for ~five min at a self-selected intensity. After a short rest period, the subject commenced the test at a workload of 100 W. This workload was maintained for 60 sec, after which the workload was increased by 50 W for a further one min. Thereafter, the workload was increased by 25 $W.min^{-1}$ until volitional fatigue. PPO was defined as the last completed workload (W) plus the fraction of time spent on the final, uncompleted workload, multiplied by the 25 W

workload increase (Hawley and Noakes, 1992). During the maximal tests, subjects wore a mask covering the nose and mouth. Expired air was passed through an on-line computer system attached to an Oxycon Alpha automated gas analyser (Mijnhardt, The Netherlands) for the determination of oxygen consumption. Before each test, the gas analyser was calibrated with a Hans Rudolph 5530, 3 L syringe and a 5% CO_2:95% N_2 gas mixture. Analyser outputs were processed by an IBM computer, which calculated minute ventilation, carbon dioxide production and oxygen consumption using conventional equations (Jones, 1982). Each subject's VO_{2peak} was taken as the highest oxygen uptake measured during any 60-sec period of the maximal test (Noakes et al., 1990).

Exactly three min after completion of the maximal test, a venous blood sample (~5 ml) was drawn directly from an antecubital vein of the forearm for the subsequent determination plasma lactate concentration. The blood sample was placed into tubes containing potassium oxalate and sodium fluoride. The blood samples were kept on ice until centrifuged at 3,000 $rev.min^{-1}$ for 10 min at 4°C, and the supernatant stored at -20°C for later analysis. Plasma lactate concentrations were determined by spectrophotometric (Beckman Spectrophotometer - M35) enzymatic assays (Lactate PAP, bioMerieux Kit, Marcey l'Etoile, France). The CV for this assay in our laboratory is <3% for duplicate lactate samples.

4.2.3 Time trials

Three days after their maximal test subjects returned to the laboratory to perform the first of three 2,000-m TTs on the Concept II™ rowing ergometer. Each TT was separated by three days. TTs were conducted at the same time of day for each subject. For the two days prior to a TT, subjects maintained the same training and dietary regimen. The coach divided the

eight-man squad into pairs according to their rowing ability, and for each of the subsequent TTs these pairs "rowed against each other". The two rowing ergometers were placed next to each other so that each subject could not see their opponent's elapsed time. A fan was positioned between the ergometers to cool the subjects. In order to minimise any potential performance improvements due to changes in training status, all of the testing was performed within 14 days of each subject's first visit to the laboratory. During all testing, laboratory conditions were standardised at a temperature of 20°C and a relative humidity of 55%.

After a standardised 5-min warm up and a short rest, the coach started the TT, during which the subjects were instructed to perform the 2,000 m in "the fastest time possible". During the TT, the only feedback given to subjects was their elapsed time for each 500-m split and the distance remaining. Subjects were not informed of the overall time for the TT. Strong verbal encouragement was provided to all subjects by the coach during all testing.

Three min after each subject's second TT, a blood sample was drawn from an antecubital vein for the measurement of plasma lactate concentrations, as described previously.

During all testing, HR were recorded with a Sport Tester HR-monitor (Polar Electro OY, Kempele, Finland). This monitor consists of an electrode belt worn around the chest, a transmitter, and a wrist mounted receiver. The receiver recorded and stored the subject's HR at 5-sec intervals.

4.2.4 Statistical analyses

Reproducibility of the time taken to row 2,000 m was derived by two-way ANOVA, as described previously. Briefly, the time taken to complete a trial was the dependent variable in the ANOVA, and identities of subject and time trial were the two effects. The natural logarithm of time was used in the ANOVA to ensure residuals in the model were normally distributed, and to permit statistically elegant derivation of the mean CV. Reliability expressed as an intraclass correlation coefficient was calculated as (F -1)/(F + k - 1), where F was the F-ratio for the subject term and k (= 3) was the number of trials. The p-value for the F-ratio of the TT term represented a test of significance of differences in the mean times for the three trials. A mean CV was derived from the root mean square error (RMSE) in the analysis by the following transformation: $CV = 100(e^{RMSE} - 1) \approx 100(RMSE)$. The 95%CI for the intraclass correlation coefficient and CV were calculated by the methods of McGraw and Wong (1996) and Tate and Klett (1959) respectively. For comparison with the CV of the present study, unbiased estimates of published CVs were obtained as previously described. CVs for the individual rowers were calculated by dividing each rower's mean within-trial SD (calculated from the mean of the variances) by his mean time for the three trials. All measures of centrality and spread are presented as mean ± SD.

4.3 Results

The characteristics of the subjects under investigation, along with data measured during the maximal test, are displayed in Table 4.1.

Table 4.2 shows the physiological responses of the subjects during the TTs. The HR of the subjects during the three TTs rose rapidly during the first 150 sec to 193 min^{-1} and reached a

peak of ~202 min^{-1} during the last 60 sec of the row. HR averaged 191 ± 7 min^{-1} (95 ± 1% of HR_{peak}) throughout the duration of the trials. Figure 4.1 illustrates the average power output measured at each successive 500-m segment for the eight subjects during each TT. Subjects sustained 92 ± 5% of PPO for the duration of the three TTs (Table 4.2). Plasma lactate concentration peaked at 19.4 ± 2.0 $mmol.l^{-1}$ after completion of TT_2.

Table 4.3 indicates the average time taken for each 500-m segment of the 2,000 m during each of the TTs. The average times (mean ± SD) for TT_1, TT_2 and TT_3 were 6:56 ± 0:15, 6:54 ± 0:14 and 6:51 ± 0:14 min:sec, respectively. The individual time ± SD taken for each TT of each rower, together with the group mean ± SD for the three trials, are shown in Table 4.4.

The CV for the three TTs was 0.6 % (95%CI 0.4 to 1.0). The reliability, expressed as an intraclass correlation coefficient, was 0.97 (95%CI 0.91 to 0.99). There was a small increase in reproducibility calculated for TT_2 and TT_3 alone (CV = 0.64%, correlation = 0.97) in comparison with that of TT_1 and TT_2 alone (CV = 0.70%, correlation = 0.96).

The lactate concentrations measured after the second TT were significantly higher than those measured after the maximal test (19.4 ± 1.9 vs 15.7 ± 1.9 $mmol.l^{-1}$, $p < 0.05$).

Table 4.1 Characteristics of the subjects.

	Mean ± SD	Range
Age (yrs)	16 ± 1	15 - 17
Height (m)	1.84 ± 0.5	1.78 - 1.92
Mass (kg)	77.2 ± 7.5	69.5 - 89.5
VO_{2peak} ($ml.kg^{-1}.min^{-1}$)	60.8 ± 4.9	54.2 - 67.7
($l.min^{-1}$)	4.7 ± 0.5	4.02 - 5.35
HR_{peak} (min^{-1})	201 ± 10	184 - 213
PPO (W)	346 ± 35	300 - 400
P:W ($W.kg^{-1}$)	4.5 ± 0.3	3.9 - 4.8
Lactate ($mmol.l^{-1}$)	15.7 ± 1.9	12.0 - 18.0

VO_{2peak}, peak oxygen uptake; HR_{peak}, peak heart rate attained during maximal test; PPO, peak sustained power output; P:W, power to weight ratio (n = 8).

Table 4.2 Physiological responses measured during the 2,000 m time-trials.

	Time (min:sec)	Average Power (W)	Average HR (min^{-1})
Trial 1	6:56 ± 0:15	313 ± 35	194 ± 7
(range)	(7:09 - 6:32)	(283 - 371)	(186 - 206)
Trial 2	6:54 ± 0:14	320 ± 33	191 ± 8
(range)	(7:10 - 6:28)	(282 - 382)	(180 - 206)
Trial 3	6:51 ± 0:14	324 ± 36	188 ± 6
(range)	(7:05 - 6:29)	(290 - 380)	(178 - 193)
Mean	6:54	319	191
SD	14	35	7

Values are mean ± SD for the three time-trials (n = 8).

Table 4.3 Split times and overall time for the 2,000-m time trials.

	Distance				
	First 500 m	**Second 500 m**	**Third 500 m**	**Final 500 m**	**2000 m**
			(min:sec)		
Trial 1	1:37.8 ± 0:5	1:43.0 ± 0:4	1:47.6 ± 0:4	1:48.0 ± 0:5	6:56.5 ± 0:15
Trial 2	1:39.2 ± 0:3	1:43.0 ± 0:3	1:46.0 ± 0:4	1:45.8 ± 0:4	6:54.2 ± 0:14
Trial 3	1:40.5 ± 0:4	1:42.3 ± 0:4	1:43.9 ± 0:4	1:44.5 ± 0:4	6:51.0 ± 0:14

Values are mean ± SD for the three time-trials (n = 8).

Table 4.4 Rowing performance time (min:sec) by individual subjects in the three 2,000 m time-trials.

Subject	1	2	3	4	5	6	7	8	Mean	SD
Trial 1	7:09	7:03	6:32	7:09	6:55	7:08	6:35	7:02	6:56.5	0:15
Trial 2	7:05	7:02	6:28	6:59	6:54	7:10	6:38	6:56	6:54.2	0:14
Trial 3	7.00	-	6:29	7:01	6:51	7:05	6:32	6:59	6:51.0	0:14
Mean	7:05	7:02	6:30	7:03	6:53	7:08	6:35	6:59		
SD	0:05	0:01	0:02	0:05	0:02	0:03	0:03	0:03		
CV (%)	1.13	0.14	0.46	1.28	0.58	0.70	0.76	0.72	0.60%	

CV, coefficient of variation. Values are mean ± SD for the three time-trials (n = 8).

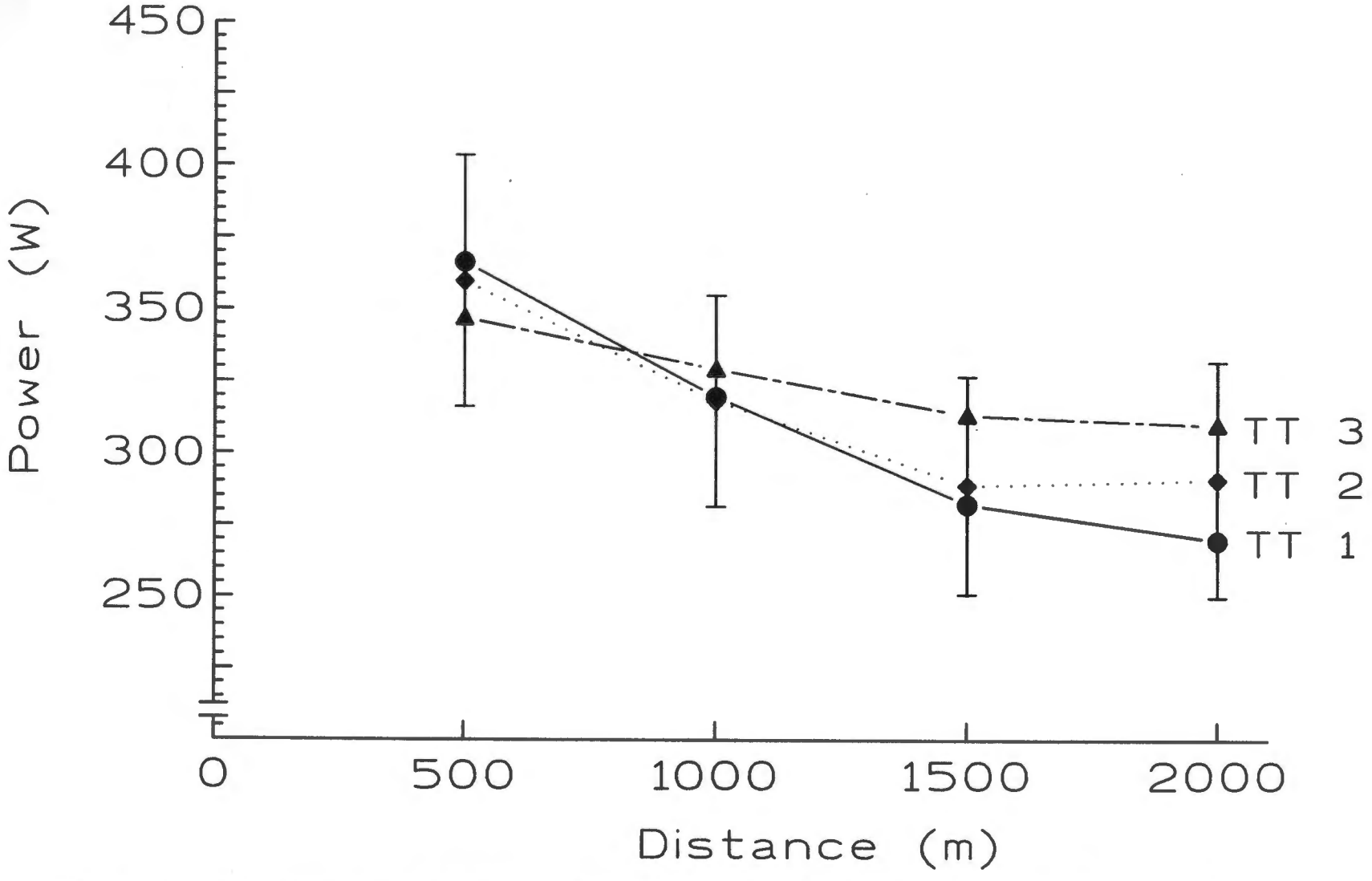


Figure 4.1 Mean power output for each successive 500-m of the 2,000-m time-trials. Values are mean ± SD.

4.4 Discussion

The main finding of the present study was that laboratory simulated performance on the Concept II™ rowing ergometer was highly reproducible. The CV for the three separate TTs was lower than that previously reported for either trained runners undertaking a self-paced treadmill task (Billat et al., 1994), or cyclists performing 20- and 40-km timed rides (Palmer et al., 1996). Such reliability is important in sports such as rowing where maximal ergometer tests are frequently employed to assess changes in the physiological status of athletes, and as a tool for team selection (Hagerman, 1994).

There are several possible explanations for the differences in the reproducibility of various laboratory measures of performance. Tests in which athletes are required to exercise to exhaustion at a fixed, submaximal workload are "open-ended", likely to be influenced by psychological factors such as motivation and boredom, and thus have poor reliability. On the other hand, when subjects are asked to complete a set distance, or undertake a given amount of work as fast as possible, they appear to be able to better regulate their effort in order to optimise their performance, which probably explains the high reliability of such tests. The type of ergometer on which the athlete is tested is also likely to impact on the reliability of any outcome measure of performance (Firth, 1981), as is familiarisation of an athlete to the testing apparatus (Hickey et al., 1992; McLellan et al., 1995).

Testing subjects over their normal race distances should also contribute to higher reproducibility. Competitive athletes are accustomed to the physiological demands of their event and employ a pacing strategy that is likely to produce the best overall time. In this regard, the standard 2,000-m distance has frequently been chosen to evaluate the metabolic

responses to competition in rowers (Hagerman et al., 1979; Hagerman and Falkel, 1987; Hagerman and Korzeniowski, 1989). In such tests rowers are accustomed to initiating an all-out effort with a fast start, during which the highest power outputs and stroke ratings are attained. The power output then drops slightly for the second and third quarter of a race, and is picked up again during the final 500-m in a sustained burst to the finish line (Hagerman, 1994).

In the current study, the split-times for the first 500-m of all three trials show that subjects rowed the first quarter of the race the fastest (Table 4.3). However, despite slightly different pacing strategies for the second and third 500-m segments of each TT, they were unable to increase their speed over the last quarter (Table 4.3). Irrespective of small differences in pacing strategies between the three trials, the overall times were not significantly different (Table 4.3). Previous studies examining the effect of different pacing strategies have mainly been conducted on cycling, running and speed-skating performance (Foster et al., 1994). In a study examining 2,000 m cycling performance, (Foster et al., 1993b), they showed that when cyclists were forced to cycle at various predetermined speeds during the first 1,000 m of a 2,000-m time-trial, the best overall performance was obtained when subjects maintained an even pacing strategy for the total distance. There was a U-shaped relationship between the relative starting pace (as a percentage of best previous time) and the finishing time. However, it is difficult to extrapolate these findings of cyclists to rowing pacing.

Of interest was the observation that the physiological responses measured during and after the simulated TT (Table 4.2) were more severe than those attained during the maximal test (Table 4.1). The peak lactate concentrations of around 19 $mmol.l^{-1}$ measured in our subjects after

the TT are similar to those reported by Hagerman et al. (1979) following an all-out 6-min rowing effort. Differences in the physiological responses between the maximal test and the simulated TTs are likely due to the fact that the 2,000-m TT closely resembles the athletes' specific competitive rowing task, whereas the incremental test used to determine VO_{2peak} does not mimic normal training or racing practices. Others (Foster et al., 1993a; Palmer et al., 1994) have previously reported that trained athletes attain higher HR in races compared to those measured during "maximal" laboratory tests.

In conclusion, the results of this study show that in trained rowers, laboratory simulated performance over a 2,000-m TT on the Concept II™ rowing ergometer was highly reproducible. Therefore, this test can be used by coaches to detect small changes in performance and sports scientists to determine the effects of nutritional and training interventions. Differences in HR_{peak} and peak lactate concentrations after the simulated competitive effort compared to the maximal test highlight the need for sports scientists to develop testing protocols which closely mimic the specific physiological demands of an athlete's competitive event.

CHAPTER FIVE

A NOVEL AND RELIABLE, LABORATORY PERFORMANCE TEST FOR ENDURANCE-TRAINED CYCLISTS

Abstract

The purpose of this study was to determine the reliability of a prolonged endurance test which mimicked racing conditions, and was conducted on the subjects' own equipment.

Eight endurance trained cyclists (VO_{2peak}, 64.8 ± 5.7 ml.kg.$^{-1}$min^{-1}; PPO, 411 ± 43 W), performed the 100-km TT on three occasions, each separated by 5-7 days. Subjects were free to regulate the intensity at which they exercised throughout the rides. Four 1-km and four 4-km sprints were included during the 100-km TTs during which the subjects were asked to cycle as fast as possible. During the trials subjects were only allowed to see their HR and the distance covered and were not informed of their time until completion of the final TT. The final times of the three TTs showed a small improvement from TT_1 to TT_3 (151:52 ± 10:36, 148:24 ± 9:42 and 147:24 ± 8:48 min:sec, for TT_1, TT_2 and TT_3, respectively). The CV for individual cyclists was 1.7% (95%CI 1.1 to 2.5) and the test-retest reliability expressed as an intraclass correlation coefficient was 0.93 (95%CI 0.79 to 0.98). The results of the study indicate that a laboratory performance test with intermittent bouts of high intensity exercise, undertaken on the subject's own bicycle, is highly reproducible. Laboratory tests in which subjects are allowed to freely choose their effort rather than having a workload being imposed on them, have proved to be more reliable than exercise tests to exhaustion conducted at a predetermined workload and will therefore allow sport scientists to be able to discriminate differences in performance associated with biological variation from real changes in performance as a consequence of an intervention.

5.1 Introduction

Over the past decade, a variety of laboratory tests have been used to assess the physiological status of an athlete, as well as to evaluate the effects of a variety of nutritional ergogenic aids and training strategies aimed at improving athletic performance. Unfortunately, the reliability of a large number of these tests has not been determined. When athletic performance is measured, every attempt should be made to exclude or control for the influence of extraneous factors which may mask the effect of the treatment (McLellan et al., 1995). In this regard, the reliability of any performance test is important when a treatment intervention is evaluated. If the reliability of a laboratory performance test is not known, the sport scientist will not be certain whether a change in performance is due to the effect of the intervention or the result of natural day-to-day variability.

Until recently, there have been few studies examining the reproducibility of performance tests in the laboratory. Previous protocols for evaluating performance have mostly relied upon the time to complete a fixed workload during cycling exercise (Bishop, 1997; Hickey et al., 1992; Jeukendrup et al., 1996; Palmer et al., 1996), the exercise time to exhaustion at a constant, predetermined workload (Boileau et al., 1977; Coggan and Costill, 1984; Jeukendrup et al., 1996; Krebs and Powers, 1989; Kuipers et al., 1985; McLellan et al., 1995) or running time to exhaustion at maximal aerobic speed (Billat et al., 1994) to assess performance. The performance measures when subjects are asked to exercise to exhaustion at a fixed intensity often have large variability (Billat et al., 1994; Krebs and Powers, 1989; McLellan et al., 1995) and, for the most part have been used to assess performance of high-intensity exercise lasting one hour or less.

Few laboratory studies have attempted to simulate field conditions. Tests in which race conditions are simulated are an important consideration because in sports such as road cycling, races are normally characterised by periods of sustained steady-state exercise punctuated by repeated bouts of high- and low-intensity work. Such alternating workbouts frequently elicit greater physiological responses than those evoked in the laboratory (Foster et al., 1993a; Palmer et al., 1996) and are in contrast to the submaximal, constant cycling intensities often used in performance tests. Regardless the stochastic nature of competitive road cycling, to my knowledge, sport scientists have not attempted to create and validate endurance cycling tests that simulate the physiological demands of road cycling races.

Therefore, the aim of the present study was to determine the reliability of an endurance cycling protocol that was designed to mimic the competitive demands of road racing, consisting of a 100-km TT with intermittent bouts of high intensity work.

5.2 Methods

5.2.1 Subjects

Eight competitive, endurance-trained cyclists and triathletes participated in this study. These subjects were chosen because they were highly motivated and accustomed to exercising for prolonged periods (2-4 hr). Prior to commencement of the trial, all subjects were informed of the nature of the investigation, after which they gave written informed consent in accordance with the guidelines outlined by the American College of Sports Medicine (1988). The subjects in this trial were well-trained. Five of the subjects had participated in previous trials in our laboratory and were familiar with laboratory procedures. The other three subjects performed a familiarisation trial after their maximal test. During this ride they performed the

first quarter of the 100-km TT (described subsequently). The characteristics of the subjects are shown in Table 5.1.

5.2.2 Preliminary testing

All subjects completed a progressive incremental test to exhaustion for the determination of VO_{2peak} and PPO on their own bicycles, which were mounted on an air-braked ergometry system (Kingcycle Ltd, High Wycombe, U.K.) described subsequently. After a 5-10 min warm up at a self selected intensity, the test commenced at a workload of 100 W after which it was progressively increased by 20 $W.min^{-1}$ until the subject could no longer maintain the required power output. The subject's PPO was taken as the highest average power during any 60-sec period of the exercise test.

During the maximal tests, subjects wore a mask covering the nose and mouth; the expired air passed through an on-line computer system attached to an Oxycon Alpha automated gas analyser (Mijnhardt, The Netherlands). Before each test, the gas analyser was calibrated with a Hans Rudolph 5530, 3 L syringe and an online CO_2:N_2 gas mixture of know composition. Analyser outputs were processed by an IBM computer which calculated oxygen uptake and carbon dioxide production using conventional equations (Jones, 1982). Each subject's VO_{2peak} was taken as the highest oxygen uptake measured during any 60-sec period of the test (Noakes et al, 1990).

During both the VO_{2peak} tests and 100-km TT, HR were recorded using a Polar Sport Tester HR-monitor (Polar Electro OY, Kempele, Finland). This monitor consists of an electrode belt worn around the chest, a transmitter, and a wrist mounted receiver. The receiver

recorded and stored the subject's HR at 5-sec intervals for the incremental test and 60-sec intervals for the TTs.

5.2.3 Kingcycle ergometry system

All testing was conducted on a Kingcycle ergometry system, which allows cyclists to ride on their own racing bicycles in the laboratory. The total resistance provided by the ergometer is equivalent to that experienced by a cyclist of approximately 65 kg riding on a flat road.

After the front wheel was removed, the subject's bicycle was attached to the ergometry system by the front fork and supported by an adjustable pillar under the bottom bracket. The bottom bracket support was used to position the rolling resistance of the rear tyre correctly on an air-braked flywheel. A photo-optic sensor monitored the velocity of the flywheel in revolutions per second (RPS), from which an IBM-compatible computer calculated the power output (W) that would be generated by a cyclist riding at that speed on a level terrain, using the following equation:

$$W = 0.000136\ RPS^3 + 1.09\ RPS$$

Before each test, the Kingcycle was calibrated by a series of "run down" calibrations. Subjects accelerated to a workload of ~300 W after which they were instructed to immediately "stop pedalling", while remaining seated in their riding position. The bottom bracket support was adjusted until the computer display indicated that the slowing of the flywheel matched a pre-determined reference power decay curve. The time taken for a laboratory simulated 20-km and 40-km time-trials on the Kingcycle ergometer system has

been shown to be highly reproducible (CV 1.1 ± 0.9% and 1.0 ± 0.5%, respectively, Palmer et al., 1996).

5.2.4 Time trials

Each subject completed three 100-km TTs, separated by a minimum of four days and a maximum of seven. Subjects performed their TTs at the same time of the day. During all trials laboratory conditions remained constant (~20 °C, 55% relative humidity). Subjects were requested to perform the same type of training for the duration of the trial and refrain from heavy physical exercise on the day prior to a TT. Subjects completed a nutritional information sheet on which they recorded their food and fluid intake for the day preceding a TT, as well as for the day on which they performed their TT. They were then instructed to repeat this dietary regimen before each trial. If subjects did not comply with the same dietary and training regimen, they were not allowed to commence a TT.

After a standardised five-min warm up, subjects commenced the 100 km-TT. They were instructed to complete the distance in "the fastest time possible". To mimic the stochastic nature of cycle road races (Palmer et al., 1994), the 100-km TT included a series of sprints during which subjects were requested to ride "as fast as possible". There were four 1-km sprints after 10, 32, 52 and 72 km, as well as four 4-km sprints after 20, 40, 60 and 80 km. Instantaneous power output was recorded at each 500-m split of both the 1-km and 4-km sprints. This provided an estimate as to the average power maintained for the whole sprint. Subjects were given a diagram before each ride showing the "course profile", and this schematic was displayed alongside the computer terminal throughout the trial. The only feedback given to subjects during TTs was their elapsed distance and HR. Subjects were not

informed of their times for any of the TTs or sprints until completion of the experiment. Throughout each trial, power output, speed and elapsed time were monitored continuously and stored by an IBM computer. At the end of each TT, the average power output per min and total time taken for the 100-km TT was downloaded from the computer. The average speed for each sprint was calculated from the time taken to complete each of these sprints. A fan was positioned to cool the subjects during their TTs, and verbal encouragement was given to each subject by the same investigator.

During the first TT, subjects were allowed access to fluid and food ad libitum. Fluids available were water and a 6.8% carbohydrate-electrolyte (184 mg sodium, 23 mg potassium per 100 ml concentrate) drink (Energade, Bromor Foods, Salt River, South Africa). The solid food consisted of bananas. The quantity of fluid and food consumed during the first trial was recorded and subjects were required to repeat the same drinking/eating regimen for the two subsequent rides.

5.2.5 Statistical analyses

Measures or reliability were derived by repeated-measures ANOVA using the mixed procedure in SAS for the Macintosh (version 6.10, SAS Institute, Cary, NC). The mixed procedure permits modelling of sources of variation, as well as the usual modelling of means for fixed effects.

For variables measured during each trial (duration, mean power, and mean heart rate for the 1-km and 4-km sprints), three independent sources of variation were modelled: between subjects, between trials, and between sprints. Means for the three levels of trial and the four

levels of sprint were included in the model as fixed effects. This full model was used to determine the statistical significance of any overall difference between means of the trials or sprints, and the statistical significance of any interaction (differences in sprints between trials); it was also used to compute values of selected differences between the means of the three levels of trial and the four levels of sprint, and their 95%CI.

For variables representing an outcome of a complete 100-km trial (duration of the trial, mean power, HR and sweat rate for the trial, and overall mean durations, powers and HRs of the 1-km and 4-km sprints), two independent sources of variation were modelled: between subjects and between trials. Means for the three levels of trial were also included in the model as a fixed effect. This reduced model is identical to the usual two-way ANOVA used in other studies of reliability (Hopkins, 1997). Here it was used to determine the correlation between trials, expressed as the intraclass correlation coefficient (between-subject variance divided by total variance). The intraclass correlation coefficient is unaffected by any difference in the means. The within-subject variation between trials was also derived from the model as an absolute standard deviation for HR, and as a CV for durations and powers. The CV was chosen for these performance variables, because within-subject variation for human performance is probably better modelled as a constant percentage of a subject's true performance rather than a constant absolute value. The CV was derived by performing the ANOVA on the natural logarithm of the variable, then transforming the within-subject standard deviation (SD) with the following formula: $CV = 100(e^{SD} - 1) \approx 100(SD)$ (Hopkins, 1997).

Confidence intervals for the intraclass correlation coefficient and the CV were calculated by the methods of McGraw and Wong (1996) and Tate and Klett (1959) respectively. CVs for the individual cyclists were calculated by dividing each cyclist's mean within-trial SD (calculated from the mean of the variances) by his mean time for the three trials. All measures of centrality and spread are presented as mean ± SD. For all tests, the level of significance was pre-set at $p < 0.05$.

5.3 Results

Table 5.2 shows the mean time and the physiological responses of the subjects during the three 100-km TTs. The average overall time taken for the 24 rides was 149:06 ± 2:14 min:sec. TT_1 (151:42 ± 10:36 min:sec) was significantly slower than TT_2 (148:24 ± 9:42 min:sec, $p = 0.02$), whereas there was no significant difference between TT_2 and TT_3 (147:24 ± 8:48 min:sec, $p > 0.5$). Similarly, the mean power output for TT_2 (258 ± 43 W) was significantly higher than the power output of TT_1 (245 ± 40 W, $p = 0.01$), but no significant difference was found between TT_2 and TT_3 (260 ± 40 W, $p = 0.5$). The average HR for the three TTs was 156 ± 11 and there were no significant differences between any of the trials. The average fluid intake of the subjects during the TTs was 2.16 ± 0.68 l (range 1.2-3.4 l), whereas their average sweat loss was 3.19 l (range 2.4-3.8 l) and sweat rate 1.29 ± 0.06 $l.hr^{-1}$. Table 5.3 shows a breakdown of the average time and power output for each successive 1-km and 4-km sprint. The time and power output for the last 1-km and 4-km sprints were significantly different compared to the first sprints ($p < 0.001$). By the final sprints, riders were, on average, 3 sec slower for the 1-km sprints (1:14 ± 0:06 to 1:17 ± 0:09 min:sec), and 10 sec slower for the 4-km sprints (5:24 ± 0:14 to 5:34 ± 0:22 min:sec). The corresponding

drop-off in power output was 66 W for the 1-km sprints (457 ± 116 to 391 ± 78 W) and 21 W for the 4-km sprints (319 ± 34 to 298 ± 45 W).

The time taken for each subject to complete the TTs, together with their individual mean ± SD and CV for the three trials, and the group mean data for the rides are shown in Table 5.4. Table 5.5 shows the within-subject variation and correlations between trials for performance variables and HR. Within-subject variation is shown as a CV (%) for times, power and sweat loss, and as a SD (min^{-1}) for HR. The CV for the total time for the 100-km of the group (n = 8) was 1.7% (95%CI 1.1 to 2.5).

Figure 5.1A displays the average power output sustained throughout the three 100-km TTs, whereas Figure 5.1B shows the subjects' HR response during the rides. The average power output sustained during the entire ride was 254 ± 8 W. There was a gradual increase in HR during the "steady-state" cycling sections of the TT. During the first 10 km, the average HR was 140 ± 16 min^{-1}, after which there was gradual increase to 155 ± 12 min^{-1} between 53 km and 60 km, and a further rise to 160 ± 12 min^{-1} between 85-95 km. During the final 5 km there was a rapid increase in HR so that during the final minute of the ride HR peaked at 165 ± 15 min^{-1}. The average HR response during the first 1-km sprint (164 ± 13 min^{-1}) was significantly lower than during the final 1-km sprint (170 ± 11 min^{-1}, $p < 0.002$). Similarly, the average HR sustained during the 4-km intervals ranged from 163 ± 10 min^{-1} for the first 4-km sprint, to 167 ± 11 min^{-1} for the final 4-km sprint. This increase was also significantly higher ($p < 0.001$). The average HR sustained for the duration of the TT was 156 ± 11 min^{-1} which corresponds to 86% of the HR_{peak}.

Table 5.1 Characteristics of the subjects.

	Mean ± SD	**Range**
Age (yrs)	26 ± 3.5	22 - 32
Height (m)	1.81 ± 0.55	1.72 - 1.87
Mass (kg)	77.6 ± 7.7	67.7 - 90
VO_{2peak} ($ml.kg^{-1}.min^{-1}$)	64.8 ± 5.7	56.3 - 72.8
($l.min^{-1}$)	5.04 ± 0.71	4.0 - 6.2
HR_{peak} (min^{-1})	181 ± 9	166 - 192
PPO (W)	411 ± 43	356 - 500
P:W ($W.kg^{-1}$)	5.3 ± 0.6	4.5 - 6.5

VO_{2peak}, peak oxygen uptake; HR_{peak}, peak heart rate attained during maximal test; PPO, peak sustained power output; P:W, power to weight ratio (n = 8).

Table 5.2 Mean time and physiological responses measured during the 100 km time-trials.

	Time (min:sec)	Average Power (W)	Average HR (min^{-1})	Sweat Loss ($l.hr^{-1}$)
Time-trial 1	151:42 ± 10:36	245 ± 40	155 ± 13	1.23 ± 0.10
(range)	(136:54 - 166:54)	(197 - 304)	(136 - 172)	(1.07 - 1.34)
Time-trial 2	148:24 ± 9:42*	258 ± 43*	157 ± 11	1.29 ± 0.10
(range)	(131:48 - 159:00)	(210 - 334)	(138 - 169)	(1.09 - 1.34)
Time-trial 3	147:24 ± 8:48*	260 ± 40*	156 ± 11	1.34 ± 0.19
(range)	(134:36 - 155:36)	(222 - 318)	(138 - 165)	(0.98 - 1.55)
Mean	149:06	254	156	1.287
SD	2:14	8	11	0.06

HR, heart rate; *Significantly different from time-trial 1 ($p < 0.05$). Values are mean ± SD for the three time-trials ($n = 8$).

Table 5.3 Time (min:sec) and mean power output attained in each 1-km and 4-km sprint during the three 100 km time-trials.

Trial	First sprint		Second sprint		Third sprint		Fourth sprint	
	Time (min:sec)	Power (W)	Time (min:sec)	Power (W)	Time (min:sec)	Power (W)	Time (min:sec)	Power (W)
Time-trial 1								
1 km	1:17 ± 0:06	440 ± 132	1:14 ± 0:05	419 ± 71	1:14 ± 0:06	394 ± 63	1:19 ± 0:07	385 ± 59
4 km	5:26 ± 0:14	312 ± 35	5:30 ± 0:12	303 ± 23	5:27 ± 0:14	302 ± 30	5:41 ± 0:25	281 ± 49
Time-trial 2								
1 km	1:13 ± 0:05	464 ± 108	1:14 ± 0:05	440 ± 95	1:16 ± 0:05	427 ± 83	1:16 ± 0:05	397 ± 92
4 km	5:22 ± 0:14	326 ± 36	5:23 ± 0:14	322 ± 31	5:25 ± 0:19	318 ± 38	5:29 ± 0:22	308 ± 44
Time-trial 3								
1 km	1:12 ± 0:05	468 ± 122	1:14 ± 0:05	446 ± 89	1:16 ± 0:07	399 ± 96	1:17 ± 0:05	391 ± 89
4 km	5:24 ± 0:14	319 ± 36	5:24 ± 0:12	318 ± 27	5:25 ± 0:13	314 ± 31	5:32 ± 0:18	304 ± 44
Average								
1 km	1:14 ± 0:06	457 ± 116	1:14 ± 0:05	435 ± 83	1:15 ± 0:06	407 ± 79	1:17 ± 0:09*	391 ± 78*
4 km	5:24 ± 0:14	319 ± 35	5:26 ± 0:12	314 ± 27	5:26 ± 0:15	311 ± 33	5:34 ± 0:22*	298 ± 45*

*Significantly different from sprint 1 ($p < 0.001$). Values are mean ± SD for the three time-trials ($n = 8$).

Table 5.4 Cycling performance time (min:sec) by individual subjects in the three 100-km time-trials.

Subject	1	2	3	4	5	6	7	8	Mean	SD
Time-trial 1	159:43	159:46	147:18	148:26	155:24	136:55	166:53	138:51	151:42	10:36
Time-trial 2	153:12	159:02	146:12	149:57	153:58	131:48	156:24	136:24	148:24	9:42
Time-trial 3	153:18	155:33	140:37	150:10	153:38	136:17	154:58	134:35	147:24	8:48
Mean	155:25	158:07	144:43	149:31	154:18	135:00	159:25	136:37		
SD	3:42	2:18	3:36	0:54	0:54	2:48	6:30	2:06		
CV (%)	2.4	1.4	2.5	0.63	0.61	2.1	4.1	1.6	1.7%	

CV, coefficient of variation. Values are mean ± SD for the three time-trials (n = 8).

Table 5.5 Within-subject variation and correlations (r) between trials for performance variables and heart rate. Within-subject variation is shown as a coefficient of variation (CV, %) for times, power output and sweat loss, and as a standard deviation (SD, min^{-1}) for heart rates.

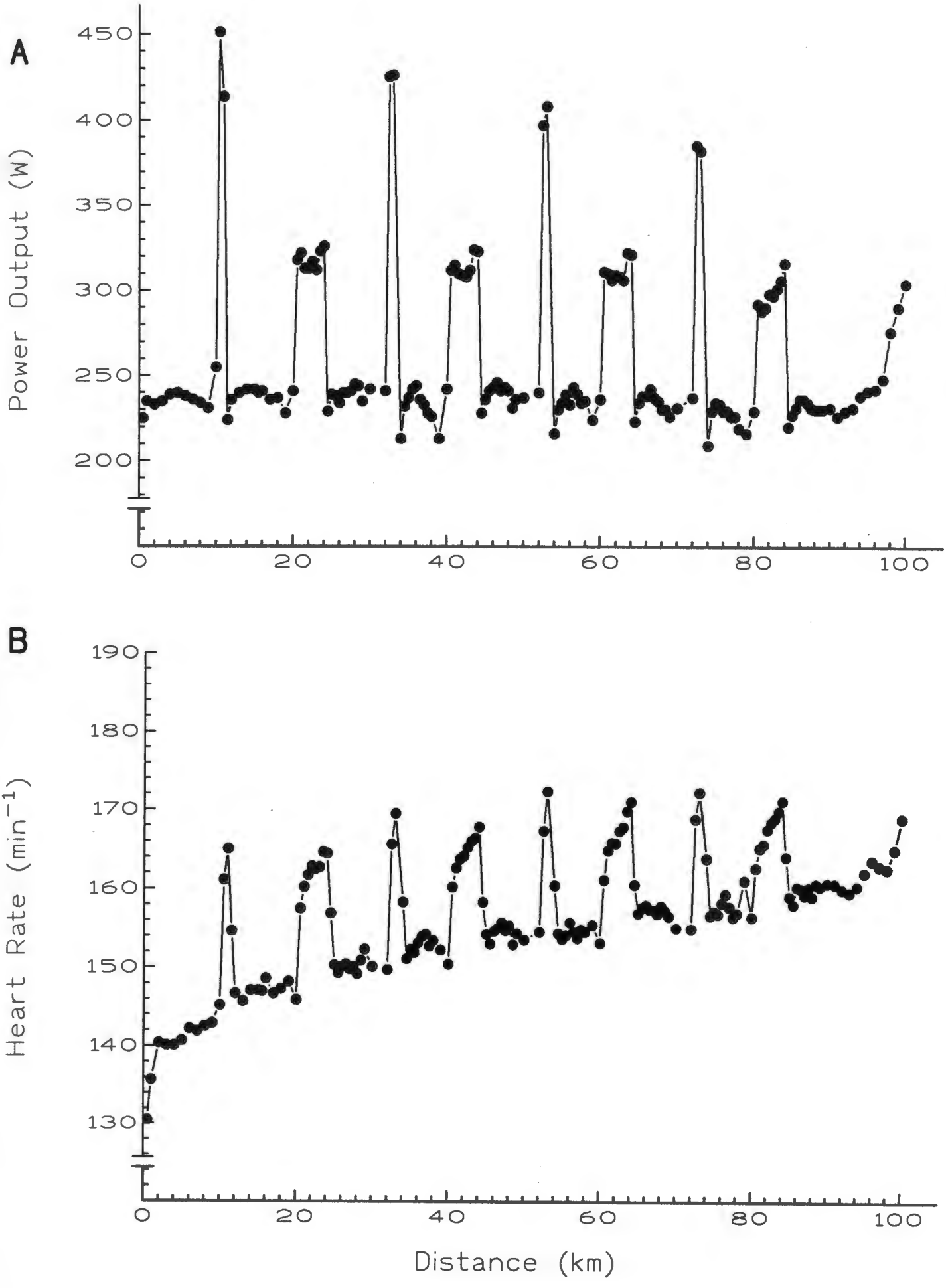
	CV (%) (95%CI)	r (95%CI)
Total time for 100-km	1.7	0.93
	(1.1 - 2.5)	(0.79 - 0.98)
Mean time for 1-km sprints	1.9	0.93
	(1.3 - 2.8)	(0.79 - 0.98)
Mean time for 4-km sprints	2.0	0.81
	(1.3 - 2.9)	(0.51 - 0.95)
Mean power for 100-km	3.7	0.94
	(2.4 - 5.4)	(0.81 - 0.99)
Mean power for 1-km sprint	4.6	0.94
	(3.0 - 6.8)	(0.81 - 0.99)
Mean power for 4-km sprint	4.9	0.81
	(3.2 - 7.2)	(0.51 - 0.95)
Mean sweat rate for 100-km	6.2	0.69
	(4.1 - 9.1)	(0.3 - 0.92)
	E (95%CI)	r (95%CI)
Mean heart rate for 100-km	3.8	0.88
	(2.5 - 5.6)	(0.66 - 0.97)
Mean heart rate for 1-km sprint	4.1	0.87
	(2.7 - 6.0)	(0.64 - 0.97)
Mean heart rate for 4-km sprint	2.6	0.93
	(1.7 - 3.8)	(0.79 - 0.98)

CV, coefficient of variation between trials for the mean 100-km TTs and of the four sprints; E, within-subject error between the mean 100-km TTs and for the mean of the four sprints; r, intraclass correlation between trials for the mean 100-km TTs and of the mean of the four sprints; 95%CI, 95% confidence interval.

Figure 5.1A Mean power output for the three 100-km time-trials.

Figure 5.1B Mean heart rate for the three 100-km time-trials.

With the exception of the 1-km and 4-km sprints, which are instantaneous power output and heart rate measurements recorded at 500 m-splits during the sprints, the data shown are averages of 1-min periods recorded during the time-trial.


A
Power Output (W)
450
400
350
300
250
200
0
20
40
60
80
100
B
Heart Rate (min^{-1})
190
180
170
160
150
140
130
0
20
40
60
80
100
Distance (km)

5.4 Discussion

The main finding of the present study was that a prolonged performance trial which included multiple bouts of high intensity exercise to simulate race conditions, was highly reproducible. Although the CV for the protocol described in the current investigation (1.7%) is slightly higher than that reported for trained cyclists undertaking 20- and 40-km TTs on a Kingcycle ergometer (CV = 1.0-1.1%, Palmer et al., 1996), it is lower than that previously reported for subjects performing a set amount of work on a cycle ergometer (Bishop, 1997; Coggan and Costill, 1984; Jeukendrup et al., 1996), or for individuals riding to exhaustion at a constant, submaximal workload (Krebs and Powers, 1989; Kuipers et al., 1985; McLellan et al., 1995). However, one should be cautious when comparing the CV reported for various test protocols. For example, even though the CV in the present study was 1.7% for the 100-km TT, there may be no difference in reliability between the CV in the present study and the lower CV of 1.1-1.2% previously reported by Palmer et al. (1996) for cyclists completing three 20-km and 40-km TTs. This is because the 95%CI found for the two different tests actually overlap each other: 95%CI of 0.7 to 1.9 and 1.1 to 2.5 for the study by Palmer et al. (1996) and the present study, respectively.

Despite the inclusion of multiple sprint bouts throughout the duration of the ride, which may have been expected to increase the variation in the overall 100-km time, trained cyclists were consistently able to reproduce their performances. This suggests that protocols during which subjects are free to chose their own effort during a simulated competitive effort are more reliable than tests in which an absolute fixed workload is imposed on them. It should, however, be mentioned that the CV of a test emphasises on the within-subject variability from test to test. Often tests to exhaustion have high correlation coefficients, despite high

CVs. For example, Billat et al. (1994) reported a CV of 25% for runners running to exhaustion at maximal aerobic speed but the test-retest correlation was 0.86. The correlations coefficient takes into consideration the between-subject differences in performance, as well as the within-subject variability.

A further factor which may contribute to the high reliability observed in this study, was that subjects were able to ride in the laboratory on their own bicycles. In previous studies when subjects were allowed to use their own equipment, high reliability has also been reported (Palmer et al., 1996).

Several physiological variables (mean power output, HR and sweat rate) were less reliable than the final performance time (Table 5.5). The higher CV observed for total power output and power output produced during the sprints compared to the CV for time could either be due to the non-linearity of the power-speed relationship, or it could have been a technical measurement error. In the present study, the higher CV for power output as opposed to the CV for time, can mainly be attributed to measurement error. As previously described (Methods, section 5.2.4), the power output during the sprints were averaged from instantaneous measurements recorded at 500-m intervals. Therefore, any changes in power output between recordings were not taken into account when calculating the "average" power output for each sprint.

As would be expected during such an intense, prolonged trial, there was a significant drop-off in power output during the high-intensity sprints (Fig. 5.1A). Although power output was significantly lower during both the final 1-km and 4-km sprints compared to the first sprint,

the decline was two-fold greater for the shorter workload (15% vs 7% for the 1-km and 4-km sprints, respectively). Despite this power decay for the sprints, the average power output during the steady-state cycling sections of the 100-km was well maintained from the start to completion of the ride (Fig. 5.1A) and even increased significantly during the final 5 km of the 100-km TT when compared to the first 5 km. It has previously been observed that subjects are able to increase their power output over the latter stages when a one-hr TT was preceded by two hrs of intense exercise (75% VO_{2max}) which included five one-min sprints at 376 W (Rauch et al., 1995).

There was a gradual increase in HR for the duration of the trial despite the relatively constant power output during the steady-state sections of the ride (Fig. 5.1B). A cardiovascular drift is often observed in trials of such a long duration even when subjects are consuming adequate amounts of fluid (Robinson et al., 1995). The average fluid intake in the current study was 0.87 $l.hr^{-1}$ which falls within the expected range of fluid intake (0.8-1.6 $l.hr^{-1}$) for cyclists during competition (Noakes, 1993). The most rapid increase in HR was observed during the final ~5 km which coincided with the increase in power output.

Despite the low overall CV for the entire 24 trials, TT_2 was significantly faster compared to TT_1. The CV represents the variation after any changes in the mean have been accounted for, which explains this apparent contradiction (see Methods). There was no substantial increase in performance between TT_2 and TT_3, indicating that there was no further learning effect. After the first 100-km TT, the subjects could have obtained a better understanding of what was expected of them and what each subject were capable of, therefore acquiring a different pacing strategy for the final two rides. Even though the subjects were familiar with

experimental procedures and had performed TTs on the Kingcycle ergometer of up to 40 km, the increase in performance between TT_1 and TT_2 confirms the need for familiarisation in tests of such a long duration or which differ from the frequently used laboratory tests. Individual CVs for seven of the subjects ranged from 0.61% to 2.5%, with only one subject having a much higher CV of 4.1%. Even though this subject was familiar with laboratory conditions, the high CV might have been due to his inability to know how to pace himself over such a long duration on the first occasion. A value of around 4% might also be expected by chance.

The possibility that the improved performances in the last two TTs could have been due to a training effect is also unlikely, since the subjects were already well-trained and their current training was stable. The number of rest days between trials was also kept to a minimum in an attempt to prevent any training effect. In general, the reliability of any laboratory test is likely to decrease as the time between successive tests increases. This is an important factor to consider when planning the time-course of any experiment which requires the sport scientist to measure performance before and after a particular treatment or intervention.

In conclusion, the results of the present study indicates that well-trained cyclists are able to reproduce their race-times during a prolonged, laboratory simulated performance ride despite the inclusion of multiple bouts of high intensity work interspaced at regular intervals throughout the trial. Laboratory protocols in which subjects are allowed to freely choose their effort are more reliable than exercise tests to exhaustion at a predetermined, constant workload and will result in a more accurate measure of performance.

CHAPTER SIX

SUMMARY AND CONCLUSIONS

The experiments described in this thesis have focused on the reliability of several laboratory tests of performance, as well as identifying some of the variables that might influence the reproducibility of such measures. The reliability of a performance test should be taken into consideration when interpreting the results of any treatment intervention.

Historically, the most frequently used tests in sports science research have been protocols during which subjects have been asked to exercise at a fixed, submaximal workload until exhaustion. These "open-ended" tests, however, have shown to have poor reproducibility. In addition, many of these tests are not related to race conditions. In contrast, exercise tests over a predetermined, fixed distance, or specified time, and in which subjects are free to chose their own workrate, are more reliable.

The experiment described in Chapter Three showed that when well-trained athletes were asked to run for a given time, and were free to regulate their own pace, their performance was reproducible to within two to four percent. The reliability of this test is of the same order of magnitude as other tests of much shorter duration but is still much lower than has been previously reported for prolonged exercise tests to exhaustion. However, higher reproducibility is required for experimental studies aimed at detecting the smallest worthwhile changes in performance with small sample sizes

In Chapter Four, the reproducibility of performance over a distance which rowers race during competition, was evaluated. That study showed that highly trained rowers were able to reproduce their competitive performances to within one percent (CV = 0.6%, 95%CI 0.4 to

1.0). No other test protocol reported in the literature has been shown to have such high reliability.

Finally, in Chapter Five, a more prolonged protocol which included intermittent bouts of high intensity exercise, was evaluated. This test was chosen to mimic conditions in mass-start cycle road races. In addition, instead of riding conventional ergometers, cyclists were able to ride their own bicycles in the laboratory setting. The results of this study showed that despite the addition of numerous bouts of high intensity work and overall duration of exercise of 2-3 hrs, well-trained cyclists were able to reproduce their performance to within two percent. This reliability compares well with the reliability of shorter laboratory performance tests, as well as being much improved compared to other tests of longer duration, such as the frequently-used tests to exhaustion at fixed workloads.

It should, however, be pointed out that despite the high CV found for many exercise protocols which require subjects to exercise to volitional fatigue, some of these test protocols still have high test-retest correlations coefficients and are likely to be quite reproducible within a given population. Whereas the CV takes the test-to-test within-subject variability into account, the correlation coefficient also takes the between-subject differences in performance into account. Bigger differences between subjects will result in higher correlations.

In conclusion, the results of the experiments conducted for this thesis suggest that when evaluating a nutritional, training or other intervention, sport scientists should employ laboratory tests of performance which have high reproducibility. Such reliability can be improved if several factors are considered. Firstly, the test should be sport specific.

Secondly, athletes should ideally be allowed to utilise their own specialised equipment which they would normally use during training or a competition. Thirdly, tests should preferably be conducted over the same distance or for approximately the same duration of time as the athlete's specialised event. Finally, no fixed workload should be imposed on an athlete during a test. At all times, athletes must be allowed to self-select their own pacing strategy. When all of these criteria are met, laboratory tests of performance can be reproduced to within one to two percent. Such high reliability is essential if sport scientists are to detect the small, but worthwhile changes in performance that often separate the top athlete from his or her rivals.

CHAPTER SEVEN

REFERENCES

American College of Sports Medicine: Policy statement regarding the use of human subjects and informed consent. (1988) *Medicine and Science in Sport and Exercise* 20(v).

Armstrong LE and Costill DL. (1985) Variability of Respiration and Metabolism: Responses to Submaximal cycling and running. *Research Quarterly* 56(2): 93-96.

Astrand PO. (1952) Experimental Studies of Physical Work Capacity in Relation to Sex and Age. Copenhagen: Munksgaard.

Bar-Or O. (1981) Le test anarobie de Wingate. *Symbioses* 13: 157-172.

Bar-Or O. (1987) The wingate anaerobic test. An update on methodology, reliability and validity. *Sports Medicine* 4: 381-394.

Billat V, Renoux JC, Pinoteau J, Petit B and Koralsztein JP. (1994) Reproducibility of running time to exhaustion at VO_{2max} in subelite runners. *Medicine and Science in Sports and Exercise* 26: 254-257.

Bishop D. (1997) Reliability of a 1-h endurance performance test in trained female cyclists. *Medicine and Science in Sports and Exercise* 29(4): 554-559.

Boileau RA, Bonen A, Heyward VH and Massey BH. (1977) Maximal aerobic capacity on the treadmill and bicycle ergometer of boys 11-14 years of age. *Journal of Sports Medicine* 17: 153-162.

Caine MP and McConnell. (1995) The reproducibility of cycling to volitional fatigue in non-cyclists. *Journal of Physiology*, 489: 36P.

Coggan AR and Costill DL. (1984) Biological and technological variability of three anaerobic ergometer tests. *International Journal of Sports Medicine* 5: 142-145.

Cohen J. (1988) Statistical power analysis for the behavioural sciences (2nd ed.). New Jersey, Lawrence Erlbaum, p. 37.

Costill DL, Kovaleski J, Porter D, Kirwan J, Fielding R and King D. (1985) Energy expenditure during front crawl swimming: Predicting success in middle distance events. *International Journal of Sports Medicine* 6: 266-270.

Dotan R and Bar-Or O. (1980) Climatic heat stress and performance in the Wingate anaerobic test. *European Journal of Applied Physiology* 44: 237-243.

Evans JA and Quinney HA. (1981) Determination of resistance settings for anaerobic power testing. *Canadian Journal of Applied Sports Science* 6: 53-56.

Farrell PA, Wilmore JH, Coyle EF, Billing JE and Costill DL. (1979) Plasma lactate accumulation and distance running performance. *Medicine and Science in Sports and Exercise* 11: 338-344.

Firth MS. (1981) A sport-specific training and testing device for racing cyclists. *Ergonomics* 24: 56-571.

Foster C, Green MA, Snyder AC and Thompson NN. (1993a) Physiological responses during simulated competition. *Medicine and Science in Sport and Exercise* 25: 877-882.

Foster C, Snyder AC, Thompson NN, Green MA, Foley M and Schrager M. (1993b) Effect of pacing strategy on cycle time trial performance. *Medicine and Science in Sport and Exercise* 25: 383-388.

Foster C, Schrager M, Snyder AC and Thompson NN. (1994) Pacing strategy and athletic performance. *Sports Medicine* 17: 77-85.

Graham TE and Andrew G. (1973) The variability of repeated measurements of oxygen debt in man following a maximal treadmill exercise. *Medicine and Science in Sports and Exercise* 5: 73-78.

Hagerman FC. (1994) Physiology and Nutrition of Rowing. In Perspectives in Exercise Science and Sports Medicine (edited by DR Lamb, HG Knuttgen and R. Murray), pp. 221-302. Indianapolis: Cooper Publishers.

Hagerman FC, Hagerman GR and Mickelson TC. (1979) Physiological profiles of elite rowers. *Physician in Sportsmedicine* 7: 74-81.

Hagerman FC and Korzeniowski K. (1989) Applied rowing ergometer testing. *FISA Colloque des Entraineurs* 19: 115-133.

Hagerman FC and Falkel JE. (1987) Training the energy systems. *American Rowing* 18: 40-43.

Harrison MH, Brown GA, Cochran LA. (1980) Maximal oxygen uptake: its measurement, application and limitations. *Aviation and Space Environment in Medicine* 51: 1123-1127.

Hawley JA and Noakes TD. (1992) Peak sustained power output predicts VO_{2max} and performance time in trained cyclists. *European Journal of Applied Physiology* 65: 79-83.

Hawley JA (1997a) Laboratory and field tests of athletic performance and potential. In Basic and Applied Science for Sports Medicine. RJ Maughan (Ed), Butterworth Heinemann Publishers, Oxford (in press).

Hawley JA, Palmer GS and Noakes TD. (1997b) Effects of 3 days of carbohydrate supplementation on muscle glycogen content and utilisation during a 1-h cycling performance. *European Journal of Applied Physiology* 75: 407-412.

Henry FM. (1959) Reliability, measurement error, and intra-individual differences. *Research Quarterly* 30: 21-24.

Henry JC, Clark RR, McCabe RP and Vanderby R. (1995) An evaluation of instrumental tank rowing for objective assessment of rowing performance. *Journal of Sports Sciences* 13: 199-206.

Herbst R. (1928) Der gastoffweschel als mab der korperlichen leistungsfahigkeit. *Deutches Archives fur Klunical Medisyne* 162: 33-50.

Hickey MS, Costill DL, McConell GK, Widrick JJ and Tanaka H. (1992) Day to day variation in time trial cycling performance. *International Journal of Sports Medicine* 13: 467-470.

Hopkins WG. (1997) Reliability: Calculations and more. In: A New view of Statistics. http://www.sportsci.org/resource/stats/relycalc.html#cv

Hughson RL, Orok CJ and Stuart LE. (1984) A high velocity treadmill running test to assess endurance running potential. *International Journal of Sports Medicine* 5: 23-25.

Jeukendrup A, Saris WHM, Brouns F and Kester ADM. (1996) A new validated endurance performance test. *Medicine and Science in Sports and Exercise* 28: 266-270.

Jones NL. (1982) Clinical exercise testing. Third edition. W. B. Saunders CO., Philadelphia. Appendix B. p. 292-300.

Katch VL, Sady SS and Freedson P. (1982) Biological variability in maximum aerobic power. *Medicine and Science in Sports and Exercise* 14: 21-25.

Kolbe T, Dennis SC, Selley E, Noakes TD and Lambert MI. (1995) The relationship between critical power and running performance. *Journal of Sports Sciences* 13: 265-269.

Krebs PS and Powers SK. (1989) Reliability of laboratory endurance tests. *Medicine and Science in Sports and Exercise* 21: S10.

Kuipers H, Verstappen FTJ, Keizer HA, Geurten P and van Kranenburg G. (1985) Variability of aerobic performance in the laboratory and its physiologic correlates. *International Journal of Sports Medicine* 6: 197-201.

Kyle SB, Smoak BL, Douglass LW and Deuster PA. (1989) Variability of responses across training levels to maximal treadmill exercise. *Journal of Applied Physiology* 67(1): 160-165.

Léger LA, Seliger V and Brassard L. (1980) Backward extrapolation of VO_{2max} values from the O_2 recovery curve. *Medicine and Science in Sports and Exercise* 12(1): 24-27.

Liljestrand G and Stenstrom N. (1920) Studien uber die physiologie des schwimmens. *Skan. Arch. fur Physiologie* 39: 1-63.

Lindsay FH, Hawley JA, Myburgh KH, Schomer HH, Noakes TD and Dennis SC. (1996) Improved athletic performance in highly trained cyclists after interval training. *Medicine and Science in Sports and Exercise* 28(11): 1427-1434.

MacDougall JD, Wenger HA and Green HJ. (1991) Physiological Testing of the High Performance Athlete. Human Kinetics Books, Champaign, Illinois, pp. 1-6.

McArdle WD, Katch FI, Pechar GS, Jacobson L and Ruck S. (1972) Reliability and interrelationships between maximal oxygen intake, physical work capacity and step-test scores in college women. *Medicine and Science in Sports and Exercise* 4(4): 182-186.

McConnell TR. (1988) Practical considerations in the testing of VO_{2max} in runners. *Sports Medicine* 5: 57-68.

McGraw KO and Wong SP. (1996) Forming inferences about some intraclass-correlation coefficients. *Psychological Methods* 1: 30-46.

McLellan TM, Cheung SS and Jacobs I. (1995) Variability of time to exhaustion during submaximal exercise. *Canadian Journal of Applied Physiology* 20: 39-51.

Noakes TD. (1988) Implications of exercise testing for prediction of athletic performance: a contemporary perspective. *Medicine and Science in Sports and Exercise* 20: 319-330.

Noakes TD, Myburgh KH and Scall R. (1990) Peak treadmill running velocity during the VO_{2max} test predicts running performance. *Journal of Sports Sciences* 8: 35-45.

Noakes TD. (1993) Fluid replacement during exercise. *Exercise and Sport Sciences Reviews* 21: 297-330.

Nummela A. (1996) A new laboratory test method for estimating anaerobic performance characteristics with special reference to sprint running. Harri Suominen (Ed) pp. 59-60, Jyväskylä.

Padilla S, Mujika I, Cuesta G, Polo JM and Chatard JC. (1996) Validity of a velodrome test for competitive road cyclists. *European Journal of Applied Physiology* 73: 446-451.

Palmer GS, Dennis SC, Noakes TD and Hawley JA. (1996) Assessment of the reproducibility of performance testing on an air-braked cycle ergometer. *International Journal of Sports Medicine* 17: 293-298.

Palmer GS, Hawley JA, Dennis SC and Noakes TD. (1994) Heart rate responses during a 4-d cycle stage race. *Medicine and Science in Sports and Exercise* 26: 1278-1283.

Patton JF, Murphy MM and Frederick FA. (1985) Maximal power outputs during the Wingate anaerobic test. *International Journal of Sports Medicine* 6: 82-85.

Rauch LHG, Rodger I, Wilson GR, Belonje JD, Dennis SC, Noakes TD and Hawley JA. (1995) The effects of carbohydrate loading on muscle glycogen content and cycling performance. *International Journal of Sports Nutrition* 5: 25-36.

Robinson S, Edwards HT and Dill DB. (1937) New records in human power. *Science* 85: 409-410.

Robinson TA, Hawley JA, Palmer GS, Wilson GR, Gray DA, Noakes TD and Dennis SC. (1995) Water ingestion does not improve 1-h cycling performance in moderate ambient temperatures. *European Journal of Applied Physiology* 71: 153-160.

Shephard RJ. (1984) Tests of maximum oxygen intake - A critical review. *Sports Medicine* 1: 99-124.

Tate RF and Klett GW. (1959) Optimal confidence intervals for the variance of a normal distribution. *Journal of American Statistical Association* 54: 674-682.

Taylor C. (1944) Some properties of maximal and submaximal exercise with reference to physiological variation and the measurement of exercise tolerance. *American Journal of Physiology* 142: 200-212.

Taylor HL, Buskirk, ER and Henschel A. (1955) Maximal oxygen intake as an objective measure of the cardio-respiratory performance. *Journal of Applied Physiology* 8: 73-80.

Weber KT and Janicki JS. (1986) Cardiopulmonary exercise testing: physiologic principles and clinical applications. Philadelphia: W.B. Saunders.

Wright GR. (1978) Variance of direct and indirect measurements of aerobic power. *Journal of Sports Medicine* 18: 33-42.

Wyndham CH, Strydom NB, Maritz JS, Morrison JF, Peter J and Potgieter ZU. (1959) Maximum oxygen intake and maximum heart rate during strenuous work. *Journal of Applied Physiology* 14: 927-936.